



Università  
di Catania

UNIVERSITÀ DEGLI STUDI DI CATANIA  
DIPARTIMENTO DI GIURISPRUDENZA  
DOTTORATO DI RICERCA IN GIURISPRUDENZA  
XXXIV CICLO

---

---

TESI DI DOTTORATO  
di MARTA GIUCA

---

**INTELLIGENZA ARTIFICIALE E RESPONSABILITÀ PENALE**

Tutor: Chiar.ma Prof.ssa Anna Maria Maugeri  
Coordinatore: Chiar.ma Prof.ssa Anna Maria Maugeri

---

---

A N N O A C C A D E M I C O 2 0 2 0 / 2 0 2 1



## Sommario

INTRODUZIONE.....	8
-------------------	---

### CAPITOLO I

INTRODUZIONE SUI SISTEMI DI INTELLIGENZA ARTIFICIALE.....	11
---	----

1. “Can machines think?” .....	11
2. “Basta aggiungere dati”. Breve spiegazione del <i>machine learning</i> da parte di un giurista .....	19
2.1 <i>Tasks</i> : quali compiti possono essere affidati a una macchina ‘che apprende’ ....	21
2.2 <i>Performance measure</i> : la valutazione delle prestazioni dell’algoritmo .....	23
2.2.1 La valutazione dell’algoritmo che procede alla classificazione .....	24
2.2.1.1. Accuratezza .....	24
2.2.1.2 Matrice di confusione .....	25
2.2.1.3 Oltre l’accuratezza: precisione, sensibilità, specificità, <i>fall-out</i> .....	27
2.2.2 Sbagliando si impara. <i>Bias</i> e varianza .....	31
2.2.3 <i>Underfitting</i> e <i>overfitting</i> . Come cucire l’abito perfetto .....	32
2.3 <i>Experience</i> : in che modo può apprendere un algoritmo .....	36
2.3.1 <i>Supervised vs unsupervised learning</i> . L’apprendimento per rinforzo .....	36
3. Modelli di apprendimento. Simbolisti e connessionisti, <i>traditional machine learning</i> e <i>representation learning</i> .....	38
3.1 Alberi di decisione.....	43
3.2 Reti neurali .....	46
3.2.1 <i>Perceptron</i> : come impara la rete .....	47
3.2.2 <i>Deep learning</i> : pregi e difetti .....	51
4. Dati, Big Data, <i>bias</i> . L’algoritmo discriminatorio.....	54
4.1 Dalla neutralità alla esplicabilità della decisione algoritmica.....	61
5. Una conclusione per iniziare .....	66

### CAPITOLO II

INTELLIGENZA ARTIFICIALE E DIRITTO .....	69
--	----

1. L’Intelligenza Artificiale nell’evoluzione del diritto dell’Unione Europea. L’esigenza di una IA antropocentrica .....	71
1.1 La Risoluzione del Parlamento europeo del febbraio 2017 e l’apertura verso il riconoscimento della personalità giuridica del robot.....	73
1.2. Gli interventi della Commissione europea .....	77

1.2.1 Il Gruppo di Esperti di Alto Livello e gli Orientamenti etici per una IA affidabile .....	78
1.2.2 Il Libro Bianco sull'IA .....	81
1.2.2.1 Individuazione delle aree di intervento della legislazione speciale .....	84
1.3 La definizione di IA .....	88
1.4 La proposta di regolamento del Parlamento europeo e del Consiglio dell'ottobre 2020. L'abbandono dell'idea di una personalità giuridica del robot. L'introduzione della nuova figura dell'operatore.....	92
1.5 La proposta di regolamento della Commissione europea.....	96
1.6 Ripartizione degli obblighi tra gli operatori economici coinvolti: dovere di diligenza e posizione di garanzia.....	99
1.7 L'impiego di sistemi di IA nella giustizia penale .....	101
2. Intelligenza Artificiale e diritto civile.....	102
2.1 La responsabilità del produttore .....	103
2.2 La responsabilità dell'utente.....	108
3. Responsabilità aquiliana e responsabilità penale: fatto illecito e fatto costituente reato.....	110
4. Diritto e tecnologia: regole di convivenza .....	112

### CAPITOLO III

INTELLIGENZA ARTIFICIALE E RESPONSABILITÀ PENALE. LA TENUTA DELLE CATEGORIE TRADIZIONALI DEL DIRITTO PENALE D'EVENTO DI FRONTE AI NUOVI RISCHI.....	115
---	-----

SEZIONE I PECULIARITÀ DEL FATTO TIPICO COLPOSO NEL CONTESTO DELLE MACCHINE INTELLIGENTI .....	121
---	-----

1. Il campo d'indagine: il reato colposo.....	121
2. La dimensione oggettiva della colpa .....	125
2.1 Genesi e individuazione della regola cautelare: la riconoscibilità e l'evitabilità .....	128
3. Tipicità colposa e giudizio causale a confronto .....	130
4. Principio di precauzione e cautele nelle attività sperimentali .....	133
5. Rischio consentito.....	137
6. La posizione del produttore. La perimetrazione della colpa specifica .....	144
6.1 Produttore modello e prodotto modello .....	150
6.2 Diligenza artificiale .....	152
6.2.1 Base ontologica e base nomologica della riconoscibilità artificiale .....	153
6.2.2. Negligenza artificiale .....	157

7. La posizione dell'utente. Diligenza umana e affidamento.....	165
7.1 Obblighi informativi dell'utente .....	168
7.2 Dovere di vigilanza e dovere di intervento .....	174
7.2.1 Macchine etiche? La violazione necessitata della regola di diligenza .....	182
8. Diligenze a confronto.....	185
 SEZIONE II LA CAUSALITÀ.....	 190
1. "Non una semplice coincidenza, ma una vera relazione di causalità" .....	190
Il criterio nomologico di accertamento della causalità come espressione del principio di legalità .....	190
2. Le leggi scientifiche.....	199
3. La causalità generale e l'aumento del rischio.....	206
3.1 La probabilità nelle tecnologie di IA: margine di errore e <i>accuracy</i> come strumenti a servizio della decisione giudiziaria .....	207
3.2 Indagine epidemiologica e risultati del test di validazione dei sistemi di IA: (tante) analogie e (poche) differenze .....	211
4. La causalità individuale o specifica .....	219
4.1 L'incertezza: la causalità individuale rinchiusa in una scatola nera. Opacità <i>ex post</i> e <i>explainable IA</i> .....	224
4.2 Verso il superamento dell'opacità causale?.....	228
5. Brevi riflessioni sulla causalità omissiva.....	236
5.1 Alcuni aspetti problematici. La difficile distinzione tra azione e omissione ...	240
5.2 (segue) La prova per esclusione .....	247
6. Nuove prospettive.....	250
 SEZIONE III LA COLPEVOLEZZA NEL REATO COLPOSO DEL PRODUTTORE E DELL'UTENTE DEL SISTEMA DI IA .....	 252
1. Colpevolezza colposa .....	252
1.1 L'esigibilità .....	255
2. L'accertamento della misura soggettiva della colpa .....	260
3. La colpevolezza colposa del produttore .....	262
4. La posizione dell'utente .....	266
5. Le circostanze anormali tra la colpevolezza e il concetto di azione "cosciente e volontaria" .....	267
6. Caso fortuito, forza maggiore e colpevolezza colposa.....	271
7. L'inquadramento di alcuni fattori problematici: a) l'autoapprendimento .....	277

7.1 (segue) b) Manomissione e manipolazione dell'algoritmo .....	279
7.2 (Segue) c) Cattiva informazione o informazione incompleta dell'utente.....	280
7.3 (segue) d) Il malfunzionamento improvviso e incontrollabile .....	282

#### CAPITOLO IV

#### DISCIPLINARE L'INTELLIGENZA ARTIFICIALE. LA RIFORMA FRANCESE SULLA RESPONSABILITÀ PENALE DA USO DI AUTO A GUIDA AUTONOMA

.....	288
Premessa.....	288
1. La sistematica del reato nel diritto penale francese.....	292
1.1 Il principio di legalità e lo strumento delle <i>ordonnances</i> .....	292
1.2 L'elemento materiale nei reati colposi.....	295
1.2.1 La causalità diretta tra volontà del legislatore e interpretazione della giurisprudenza.....	296
1.2.2 Il nesso di causalità indiretto .....	297
1.2.3 Certezza del nesso di causalità .....	300
1.3 La <i>faute</i> .....	302
1.3.1 L'elemento morale nelle contravvenzioni.....	305
2. Responsabilità penale e veicoli a guida autonoma. Alcune riflessioni sulla recente riforma del Code de la route.....	307
2.1 Code de la route e code pénal.....	307
3. La riforma del 2021 e la disciplina della responsabilità penale in caso di utilizzo di auto a guida autonoma.....	309
4. La responsabilità del conducente.....	312
5. La responsabilità del produttore.....	318
5.1 Il regime delle contravvenzioni. Responsabilità per il fatto del terzo? .....	324
5.2 Il dibattito sorto intorno all'art. 121-3 del code de la route e la compatibilità con i principi della responsabilità personale e della presunzione di innocenza .....	329
5.3 La lettura della norma sulla responsabilità del produttore alla luce della pronuncia del <i>Conseil constitutionnel</i> del 1999 .....	336
6. Un esempio da imitare?.....	339
RIFLESSIONI CONCLUSIVE.....	344
BIBLIOGRAFIA .....	349



## INTRODUZIONE

Quando nel corso di questi anni mi sono trovata a raccontare delle mie ricerche a chi, estraneo al mondo del diritto, mi domandava quale fosse l'argomento della mia tesi di dottorato, la reazione è spesso stata uno sguardo perplesso del mio interlocutore dopo aver sentito che il titolo da me scelto per questo lavoro fosse "Intelligenza artificiale e responsabilità penale". Allo sguardo perplesso seguiva la domanda volta a capire in che modo una tecnologia così nuova possa interessare al diritto. I dubbi del mio interlocutore svanivano non appena declinavo il mio titolo con un esempio, solitamente ricorrendo a situazioni di sinistri stradali con auto a guida assistita o automa. Vedevo allora l'espressione che mi stava davanti e rivedevo la mia quando di questo argomento ho deciso di occuparmi, la curiosità di trovare una risposta, di avere delle certezze e lo stupore, dell'interlocutore, il quale, forse per la prima volta, iniziava a considerare il diritto una scienza che stimolava anche il suo interesse. La meraviglia dell'interlocutore era anche dettata dalla convinzione che il diritto si occupa "di diritto" e non di tecnologia, almeno secondo la sua idea di persona estranea alla sfera giuridica. È, questo, un pregiudizio che in molti hanno verso noi giuristi: la convinzione che studiamo una scienza chiusa e che, perdipiù, adoperi un linguaggio troppo formale per essere compreso.

Quando io ho scelto gli studi di giurisprudenza, l'ho fatto perché già allora per me il diritto era vita, vita di relazione, di scambi, di fisiologie e patologie sociali.

In un mondo sempre più tecnologico, il "mio" diritto non poteva allora non interessarsi dei nuovi quesiti che la tecnologia pone al giurista.

Questo ha richiesto uno studio preliminare dedicato ai sistemi intelligenti. Lo studio è confluito nel CAPITOLO I, che ripercorre dapprima la nascita di queste tecnologie, per poi concentrarsi sul funzionamento dell'apprendimento automatico. Si vedrà che diversi sono i modelli di apprendimento e che si tratta comunque di un apprendimento "senza comprensione", atteso che l'algoritmo individua delle correlazioni statistiche attraverso l'elaborazione di dati, ma non è in grado di percepire il valore semantico o di avere capacità cognitive.

Il CAPITOLO II è dedicato agli interventi delle istituzioni dell'Unione Europea, principalmente del Parlamento e della Commissione, nel settore dell'IA. Si tratta di un capitolo che col tempo è andato arricchendosi. Quando ho iniziato questo studio era stata adottata la Risoluzione del Parlamento europeo del febbraio 2017, mentre della

Commissione erano presenti solo delle comunicazioni. Sono poi apparse sulla scena il documento del Gruppo di esperti di alto livello sull'IA dell'aprile 2019, il Libro bianco del febbraio 2020, la Proposta di regolamento del Parlamento europeo e del Consiglio dell'ottobre 2020 e la successiva proposta di regolamento della Commissione dell'aprile 2021. Quest'ultima ha rappresentato la base per lo sviluppo delle riflessioni in tema di tipicità colposa e dovere di diligenza gravante sul produttore e sull'utente, cui è dedicata la SEZIONE I del CAPITOLO III, con la quale il lettore è introdotto nel cuore dell'indagine, relativa alla responsabilità penale del produttore e dell'utente del sistema intelligente.

Il CAPITOLO III risulta diviso in tre SEZIONI, che si occupano del reato colposo, affrontando dapprima le questioni rilevanti in punto di illiceità colposa alla SEZIONE I, come già detto. Si passa poi allo studio della causalità, cui è dedicata la SEZIONE II del CAPITOLO III, in cui emergono temi già noti al diritto penale, ma presentati nella nuova veste del prodotto intelligente. In particolare, è individuata una simmetria tra gli studi epidemiologici e le informazioni relative all'accuratezza dell'IA, entrambi basati su generalizzazioni e, per questo, inadonei a fondare l'accertamento causale. Si dà pure atto di quegli studi finalizzati ad "aprire" la scatola nera in cui sono racchiuse le risposte relative alla scelta effettuata dell'algorithm, per consentire l'individuazione di un eventuale errore dello stesso e la conseguente riconducibilità causale a uno dei protagonisti della filiera produttiva.

Nella SEZIONE III del CAPITOLO III si affrontano le problematiche relative alla rilevanza penale di circostanze peculiari che possono presentarsi nel caso concreto. Tali circostanze vengono quindi inquadrare entro le categorie della colpevolezza colposa, del caso fortuito e della forza maggiore.

Infine, nell'ultima parte della tesi confluisce lo studio da me condotto durante il mio soggiorno all'estero presso l'Université Paris-1, Panthéon-Sorbonne la scorsa estate, quando ormai la pandemia iniziava a consentire un ritorno alla vita di relazione e di spostamenti.

Il Governo francese ha modificato il *code de la route* con l'*ordonnance 14 avril 2021 n. 2021-443*. È stata così disciplinata la responsabilità penale del conducente e del produttore di auto dotate del sistema di guida autonoma introducendo gli artt. L 123-1 c. route e L 123-2 c. route, sui quali il CAPITOLO IV si sofferma per individuarne i punti critici e interrogarsi sull'utilità della riforma.



# CAPITOLO I

## INTRODUZIONE SUI SISTEMI DI INTELLIGENZA ARTIFICIALE

1. “*Can machines think?*” - 2. “Basta aggiungere dati”. Breve spiegazione del *machine learning* da parte di un giurista - 2.1 *Tasks*: quali compiti possono essere affidati a una macchina ‘che apprende’ - 2.2 *Performance measure*: la valutazione delle prestazioni dell’algoritmo - 2.2.1 La valutazione dell’algoritmo che procede alla classificazione – 2.2.1.1 Accuratezza – 2.2.1.2 Matrice di confusione - 2.2.1.3 Oltre l’accuratezza: precisione, sensibilità, specificità, *fall-out* - 2.2.2 Sbagliando si impara. *Bias* e varianza - 2.2.3 *Underfitting* e *overfitting*. Come cucire l’abito perfetto - 2.3 *Experience*: in che modo può apprendere un algoritmo - 2.3.1 *Supervised vs unsupervised learning*. L’apprendimento per rinforzo - 3. Modelli di apprendimento. Simbolisti e connessionisti, *traditional machine learning* e *representation learning* - 3.1 Alberi di decisione - 3.2 Reti neurali - 3.2.1 *Perceptron*: come impara la rete - 3.2.2 *Deep learning*: pregi e difetti - 4. Dati, *Big Data*, *bias*. L’algoritmo discriminatorio - 4.1 Dalla neutralità alla esplicabilità della decisione algoritmica - 5. Una conclusione per iniziare.

### 1. “*Can machines think?*”

Nel noto articolo pubblicato nel 1950 sulla rivista filosofica *Mind*, Alan Turing si interrogava sulla possibilità di costruire una macchina pensante<sup>1</sup>. A tal fine, aveva elaborato un test, oggi noto come “test di Turing”, che ha la finalità di stabilire se una macchina ha un comportamento intelligente. Nell’idea di Turing, il comportamento intelligente consiste nell’abilità di raggiungere prestazioni a livello umano in tutti i compiti cognitivi (linguaggio, rappresentazione della conoscenza, ragionamento e apprendimento), sufficienti a ingannare un interrogatore<sup>2</sup>. Lo scopo era rispondere alla domanda “*can machines think?*”.

Tuttavia, tale domanda avrebbe prima necessitato che fossero definite le nozioni di “macchina” e di “pensare”. Per superare il problema definitorio, Turing elaborava un gioco: “*the imitation game*”. In breve, un giudice umano avrebbe avuto una conversazione con un interlocutore misterioso, che sarebbe stato un altro umano o una macchina. Se dalle risposte date il giudice non fosse stato in grado di capire se si stesse confrontando con una macchina ovvero con un uomo, la macchina avrebbe superato il test che l’avrebbe definita, nell’ottica di Turing, come macchina pensante.

Successivamente, l’idea che una macchina potesse simulare le caratteristiche dell’intelligenza umana ha trovato una definizione nell’espressione “intelligenza artificiale”. L’aggettivo “artificiale” denota il carattere non umano dell’intelligenza,

---

<sup>1</sup> A. Turing, *Computing, Machinery and Intelligence*, in *Mind*, LIX (236), 1950, pp. 433-460.

<sup>2</sup> S. J. Russel, P. Norvig, *Intelligenza artificiale, un approccio moderno*, Utet, Torino, 2002, p. 5 - 7.

facoltà ritenuta un tempo propria e esclusiva dell'uomo<sup>3</sup>. L'intelligenza "non umana", allora, è quella che non deriva dalla natura, ma dall'*artificium*, in quanto oggetto prodotto ad arte, attraverso un'attività creativa<sup>4</sup>.

Tale fortunata locuzione si fa risalire a John McCarthy, ricercatore di matematica presso il Dartmouth College di Hannover, New Hampshire, che fu tra gli organizzatori del convegno estivo tenutosi a Dartmouth nel 1956 sul tema.

La premessa di studio dell'argomento era che "ogni aspetto dell'apprendimento o qualsiasi altra caratteristica dell'intelligenza possono essere di principio descritti in modo così preciso che una macchina possa simularli"<sup>5</sup>.

Il Dartmouth Summer Research Project on Artificial Intelligence viene oggi considerato l'evento che diede origine all'intelligenza artificiale, in quanto la accreditò come disciplina scientifica<sup>6</sup>. Ma cosa si intendeva per intelligenza artificiale?

Nella storia della IA<sup>7</sup>, l'intelligenza è stata riferita tanto al *comportamento*, cioè alle azioni, quanto al *ragionamento*, cioè ai processi del pensiero.

Prendendo le mosse dall'idea di *comportamento intelligente*, va osservato che esso è stato descritto secondo modelli diversi, uno centrato sull'uomo, l'altro di stampo razionalista<sup>8</sup>. L'impostazione antropocentrica misura il successo in termini di prestazioni umane, mentre l'impostazione razionalista valuta la prestazione rispetto a un concetto ideale di intelligenza, definito 'razionalità'<sup>9</sup>.

Rientra nel primo ambito proprio il test di Turing, che definisce intelligente il comportamento che simuli quello umano e sia in grado di ingannare l'interlocutore. Secondo questa impostazione, l'IA è "l'arte di creare macchine che eseguono funzioni che richiedono intelligenza se vengono eseguite da persone"<sup>10</sup>.

---

<sup>3</sup> Come nota L. Floridi, il quale sostiene che stiamo assistendo a una quarta rivoluzione, si veda L. Floridi, *la quarta rivoluzione. Come l'infosfera sta trasformando il modo*, Raffaello Cortina Editore, Milano, 2017.

<sup>4</sup> A. Caselli, *Dagli artifici dell'intelligenza, all'Intelligenza Artificiale*, in *Intelligenza artificiale, protezione dei dati personali e regolazione*, F. Pizzetti (a cura di), Giappichelli, Torino, 2018, p. 192.

<sup>5</sup> J. McCarthy, M. L. Minsky, N. Rochester, C.E. Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955, p.2. Il documento è disponibile al sito <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> (ultima consultazione luglio 2020).

<sup>6</sup> G. F. Italiano, *Intelligenza artificiale: passato presente, futuro*, in *Intelligenza artificiale, protezione dei dati personali e regolazione*, F. Pizzetti (a cura di); Giappichelli, Torino, 2018, p. 209.

<sup>7</sup> Una accurata ricostruzione storica dell'evoluzione della IA è offerta da N. J. Nilsson, *The Quest for Artificial Intelligence. A History of Ideas and Achievements*, Cambridge University Press, 2010.

<sup>8</sup> S. J. Russel, P. Norvig, *Intelligenza artificiale*, cit., p. 5.

<sup>9</sup> S. J. Russel, P. Norvig, *Intelligenza artificiale*, cit., p. 4.

<sup>10</sup> R. Kurzweil, *The age of intelligent machines*, MIT Press, 1990.

Per contro, l'impostazione razionalista considera come modello non l'uomo, ma l'*agente razionale*<sup>11</sup>. Questo modello è seguito da Stuart J. Russel e Peter Norvig, i quali definiscono "agente" qualcosa che percepisce e agisce e definiscono l'IA come *progettazione di un agente razionale*<sup>12</sup>. Secondo tale approccio, "un sistema è razionale se fa la cosa giusta"<sup>13</sup>. L'idea di fondo è creare sistemi che pensano razionalmente o agiscono razionalmente, a prescindere da una loro piena equiparazione all'essere umano. L'agente razionale o intelligente consiste, pertanto, in un'astrazione<sup>14</sup>, un modello di sistema in grado di percepire l'ambiente che lo circonda attraverso dei sensori e di agire all'interno di esso attraverso degli attuatori.

Ma, come detto, il comportamento è solo un aspetto dell'*intelligenza*, la quale si compone anche dei processi del pensiero o del *ragionamento*.

Nell'ambito della IA, questo aspetto dell'intelligenza consta della *rappresentazione della conoscenza*, del *ragionamento automatico* e dell'*apprendimento automatico*<sup>15</sup>.

L'espressione 'rappresentazione della conoscenza' si riferisce alla raccolta di informazioni fornite alla macchina, quindi la questione concerne quale sistema di rappresentazione adottare per codificare la conoscenza. Il ragionamento automatico, invece, consiste nell'utilizzo delle informazioni di cui la macchina dispone al fine di assumere una decisione; l'apprendimento automatico consente invece alla macchina di adattarsi a nuove circostanze.

Va pertanto chiarito subito che l'IA non consiste in una singola tecnologia, ma in una famiglia di tecnologie, tra le quali la rappresentazione della conoscenza, il ragionamento automatico - entro il quale vanno ricomprese le aree della pianificazione e del supporto decisionale - e il *machine learning* o apprendimento automatico.

Nello studio dell'intelligenza artificiale intesa come *ragionamento*, comprendente la *rappresentazione della conoscenza*, il *ragionamento automatico* e l'*apprendimento automatico*, si sono registrati due atteggiamenti diversi da parte degli scienziati: il filone della IA di tipo logico e quello della IA di tipo biologico<sup>16</sup>.

---

<sup>11</sup> Russel, Norvig, *Intelligenza artificiale, cit.*, p. 34.

<sup>12</sup> Russel, Norvig, *Intelligenza artificiale, cit.*, p. 7

<sup>13</sup> S. J. Russel, P. Norvig, *Intelligenza artificiale, cit.*, p. 4.

<sup>14</sup> L. Carlucci Aiello, *Intelligenza artificiale*, in *Enciclopedia Italiana – IX Appendice, 2015*, [www.treccani.it](http://www.treccani.it).

<sup>15</sup> Russel, Norvig, *Intelligenza artificiale, cit.*, p. 6.

<sup>16</sup> D. Parisi, voce *Intelligenza artificiale*, in *Enciclopedia delle scienze sociali, Treccani, 1996*, in [www.treccani.it](http://www.treccani.it).

Secondo il primo approccio, per creare un sistema che “pensa razionalmente”, bisogna ricorrere alle regole della logica: la conoscenza è espressa attraverso un linguaggio formale e l’agente possiede dei mezzi per effettuare dei ragionamenti in quel linguaggio. Questo approccio è noto come *tradizione logica in intelligenza artificiale*, tra i cui sostenitori figurava lo stesso McCarthy<sup>17</sup>.

A tal fine, la maggior parte delle ricerche prendeva le mosse dal *ragionamento simbolico*, seguendo il filone simbolista della scienza cognitiva. Si tratta di “un approccio matematico in cui idee e concetti sono rappresentati da simboli, come parole, frasi o proposizioni, elaborati in seguito secondo le regole della logica”<sup>18</sup>.

L’agente che ragiona logicamente possiede un “modello a scala ridotta” della realtà esterna e delle proprie azioni possibili<sup>19</sup> e ragiona sul modo in cui agire per raggiungere un certo obiettivo.

Le rappresentazioni della realtà esterna costituiscono la base di intelligenza dell’agente e sono chiamate formule. Ogni singola formula è compresa dall’agente in quanto espressa in un linguaggio di rappresentazione della conoscenza<sup>20</sup>.

Un settore nel quale il ragionamento logico-formale nel campo dell’IA ha trovato applicazione è il gioco degli scacchi, poiché si tratta di un sistema di regole precise e definite, che consente di determinare se una data conclusione deriva dalle premesse iniziali<sup>21</sup>.

Tuttavia l’approccio logico-formale presenta non poche criticità.

Anzitutto, non tutti i problemi trovano soluzione all’interno di un sistema definito di regole formali, poiché molte situazioni si verificano in campi in cui la conoscenza è incerta<sup>22</sup>. Tradurre in linguaggio formale la conoscenza risulta difficoltoso quando

---

<sup>17</sup> Questa impostazione è sostenuta da J. McCarthy nel suo articolo *Programs with Common Sense*, 1958, disponibile online: <http://jmc.stanford.edu/articles/mcc59/mcc59.pdf> (ultima consultazione settembre 2020), oltre che da R. S. Egelmore, *Artificial Intelligence and Knowledge Based Systems: origins, methods and opportunities for NDE*, disponibile al link <https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1121&context=qnde> (ultima consultazione settembre 2020)

<sup>18</sup> AA.VV., *Macchine che pensano*, cit., p. 22.

<sup>19</sup> K. Craik, *The Nature of Explanation*, Cambridge University Press, 1943.

<sup>20</sup> Agli agenti che ragionano logicamente è dedicato il cap. 6 dell’opera di Russel e Norvig, cui si rinvia.

<sup>21</sup> A. Carobene, voce *Intelligenza Artificiale, Approfondimento*, in *Enc. Treccani online*, [www.treccani.it](http://www.treccani.it). In particolare, il modello utilizzato è quello dell’albero di decisione. L’albero del gioco indica le sequenze di tutte le mosse convenienti. Nella radice è contenuta la posizione attuale sulla scacchiera, mentre la foglia è la posizione ritenuta più vantaggiosa. La macchina costruisce l’algoritmo consistente nel percorso che conduce dalla radice alla foglia e esegue l’algoritmo facendo la mossa prescelta. Cfr. M. Somalvico, F. Amigoni, V. Schiaffonati, *La grande scienza. Intelligenza artificiale*, in *Storia della scienza, 2003*, *Enciclopedia Treccani online*, [www.treccani.it](http://www.treccani.it).

<sup>22</sup> A. Carobene, cit.

quest'ultima presenta un livello di certezza inferiore al 100%, posto che il ragionamento logico genera nuove formule necessariamente vere da formule che già erano vere, dando luogo a un'inferenza logica o deduzione.

Per questi motivi, il ragionamento tramite inferenze logiche è solo uno dei modi per raggiungere la razionalità. Accanto a esso si è sviluppato il filone biologico della IA, che mira a creare sistemi che riproducono le funzioni del cervello, simulando il funzionamento dei neuroni. Si tratta delle reti neurali o *neural networks*, che, come avremo modo di approfondire di seguito, concretizzano il pensiero connessionista della scienza cognitiva. Caratteristica del modello delle reti neurali è che esso riesce a intervenire in situazioni in cui la conoscenza è incerta, poiché caratterizzata da una molteplicità di variabili cui si associano valori non binari, non rispondenti alla logica vero-falso<sup>23</sup>.

Osservando lo sviluppo degli studi nel campo della IA, si nota che l'interesse verso la materia non è cresciuto gradualmente negli anni, ma ha conosciuto una serie di cicli, poiché a periodi di grande ottimismo e di investimenti economici sono seguiti periodi di delusione e frenata, noti come gli “inverni digitali”.

Il primo periodo, che va dal 1956 al 1974, fu quello del grande entusiasmo. A riscuotere successo fu, tra le altre, l'area dell'elaborazione del linguaggio naturale. Nel 1966 fu presentato al grande pubblico il primo *chatbot*, chiamato Eliza, da Eliza Doolittle, protagonista femminile del *Pigmaliione* di George Bernard Shaw, a cui viene insegnato il linguaggio dell'alta società inglese. L'invenzione era dell'informatico Joseph Weizenbaum e simulava uno psicoterapeuta in grado di conversare con i suoi pazienti<sup>24</sup>. Tuttavia, si trattò di una “falsa partenza”<sup>25</sup>, poiché seguì il primo “inverno digitale”.

Le ragioni del venir meno della fiducia rispetto a questo ambito della ricerca vanno rintracciate nelle tecniche utilizzate dai ricercatori in quel periodo. Degli approcci sopra descritti, inizialmente fu prediletta l'impostazione razionale secondo le regole della logica matematica. Tuttavia, presto si comprese che la creazione di una singola intelligenza artificiale equivalente a quella umana, lavorando attorno al concetto del comportamento intelligente inteso come un'astrazione, era difficile da raggiungere. I principali sistemi di

---

<sup>23</sup> A. Carobene, voce *Intelligenza artificiale*, cit.

<sup>24</sup> AA.VV., *Macchine che pensano. La nuova era dell'intelligenza artificiale*, New scientist, Edizioni Dedalo, Bari, 2018, p. 19.

<sup>25</sup> AA.VV., *Macchine che pensano*, cit., p. 18.

IA prodotti in questo primo periodo erano relativi a mondi semplificati o autosufficienti, come i giochi<sup>26</sup> e per questo venivano considerati come dei “giocattoli”.

Quanto al filone biologico, nel 1958 fu presentato da Frank Rosenblatt il perceptrone<sup>27</sup>, che avrebbe consentito più tardi, negli anni ottanta, lo sviluppo delle reti neurali. Tuttavia, il modello apparve troppo semplice e non superò le critiche mosse da studiosi matematici, in particolare quelle di Marvin Minsky<sup>28</sup>.

Ma a incidere in senso negativo sulla ricerca furono soprattutto il limitato potere di calcolo delle tecnologie disponibili negli anni '70 e la carenza di grandi quantità di dati, necessari per garantire un migliore funzionamento di molte applicazioni di IA<sup>29</sup>. Al settore dell'IA non furono più destinati i finanziamenti stanziati in passato anche da enti pubblici, come la *Defence Advanced Research Projects Agency* (DARPA) del Dipartimento della Difesa degli Stati Uniti<sup>30</sup>.

Questo primo “inverno digitale” durò fino alla fine degli anni '70. Seguì un altro ciclo di entusiasmo e di successiva delusione, con un secondo “inverno digitale” tra il 1987 e il 1993.

A partire dalla seconda metà degli anni '90, inizia una nuova stagione di successi, che durerà fino ai nostri giorni.

Nel 1997, il sistema *Deep Blue*, ideato dall'azienda IBM, batte il campione mondiale di scacchi Garry Kasparov; nel 1999 il sistema di IA chiamato *Remote Agent* ha per due giorni il controllo principale della sonda della NASA *Deep Space 1*, collocata a 100 Km dalla Terra. Si concretizza l'idea dei veicoli a guida autonoma.

Nel 2011 il sistema della IBM chiamato *Watson* vince il quiz Jeopardy battendone i campioni e allo stesso modo, nel 2017, il programma di Google chiamato *AlphaGo* batte i campioni del gioco *Go*.

È solo l'inizio di una nuova primavera dell'IA, che oggi è entrata a far parte del nostro quotidiano.

Le ragioni di questa rinascita vanno rintracciate nel progresso che settori correlati all'IA hanno conosciuto in questi ultimi decenni. In particolare è il settore di raccolta dei dati ad aver avuto una notevole espansione. Così è iniziata una nuova fase della ricerca

---

<sup>26</sup> J. Kaplan, *Intelligenza artificiale. Guida al futuro prossimo*, LUISS University Press, 2017, p. 43.

<sup>27</sup> F. Rosenblatt, *The Perceptron: a probabilistic model for information storage and organization in the brain*, pubblicato in *Psychological Review*, Vol. 65, n. 6, nel 1958.

<sup>28</sup> M. Minsky, S. Papert, *Perceptrons. An Introduction to Computational Geometry*, Mit Press, 1969.

<sup>29</sup> G. F. Italiano, *Intelligenza artificiale, cit.*, p. 213.

<sup>30</sup> J. Kaplan, *Intelligenza artificiale. Guida al futuro prossimo*, Luiss University Press, 2017, p. 44.

sui sistemi di IA: è l'età dell'intelligenza artificiale basata sui dati, che adopera una tecnologia il cui linguaggio non è solo quello logico, ma anche quello statistico<sup>31</sup>.

Tale fattore ha contribuito a direzionare la ricerca verso una specifica tecnica tra quelle comprese entro la più ampia categoria di IA. Tale tecnica è l'apprendimento automatico o *machine learning*<sup>32</sup>, studiata soprattutto a partire dagli anni ottanta, mentre era stata trascurata nel ventennio precedente. Attraverso l'apprendimento automatico si è in grado di rendere il sistema di IA autonomo, cioè di far in modo che esso proceda da solo e vada oltre le conoscenze inserite dal programmatore.

Come avremo modo di analizzare più avanti, l'apprendimento automatico può avvenire attraverso inferenze deduttive su rappresentazioni simboliche ovvero tramite tecniche di apprendimento statistico.

Negli ultimi tempi ha riscosso particolare successo il *machine learning* basato su reti neurali<sup>33</sup>, che rinuncia a riprodurre il ragionamento umano e rientra nella IA di tipo biologico poiché presta attenzione alle basi biologiche dell'intelligenza naturale<sup>34</sup>. L'approccio è subsimbolico, poiché il sistema di IA non ha una rappresentazione dettagliata della conoscenza<sup>35</sup>. L'uomo non fornisce alla macchina tutta la conoscenza, bensì un metodo di apprendimento, da applicare ai dati cui la macchina ha accesso<sup>36</sup>.

Naturalmente, il successo delle tecniche di *machine learning* che seguono un approccio diverso da quello logico va ricercato nei dati, di cui oggi esiste una quantità vasta al punto tale da essere stata coniata l'espressione *Big data*. Ciò ha favorito un intreccio dell'IA e dell'apprendimento automatico con la statistica<sup>37</sup>.

La macchina, allora, non ragiona, non pensa. "*Can machines think?*": no<sup>38</sup>. La macchina calcola e giunge a una soluzione attraverso un metodo statistico, senza

---

<sup>31</sup> AA.VV. *Macchine che pensano, cit.*, p.34.

<sup>32</sup> D. Parisi, voce *Intelligenza artificiale*, in *Enciclopedia delle scienze sociali*, Treccani, 1996, in [www.treccani.it](http://www.treccani.it).

<sup>33</sup> *Proposte per una strategia italiana per l'intelligenza artificiale*, a cura del Gruppo di esperti MISE sull'IA, Ministero dello sviluppo economico, p. 11. Il documento è disponibile online.

<sup>34</sup> D. Parisi, voce *Intelligenza artificiale*, cit., [www.treccani.it](http://www.treccani.it)

<sup>35</sup> M. Somalvico, F. Amigoni, V. Schiaffonati, *La grande scienza. Intelligenza artificiale*, cit., [www.treccani.it](http://www.treccani.it).

<sup>36</sup> F. Lagioia, G. Sartor, *L'intelligenza artificiale per i diritti dei cittadini: il progetto Claudette*, in *Ragion Pratica*, I, giugno 2020, p. 91.

<sup>37</sup> L. Carlucci Aiello, *Intelligenza artificiale*, cit., [www.treccani.it](http://www.treccani.it).

<sup>38</sup> Ma è lo stesso Alan Turing, come già detto in apertura, a riformulare la domanda e a affermare che il quesito "can machines think" fosse privo di un vero significato. Così si esprime: "*The original question, "Can machines think?" I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted*". Inoltre, sebbene minoritaria, v'è una parte di esperti che pensa che l'IA abbia una forma di coscienza. Si vedano sul punto

“comprendere” i dati, poiché la macchina si presenta come un sistema sintattico e non semantico<sup>39</sup>. Si assiste, pertanto, a uno scollamento tra il *pensare razionalmente*, che è attività lasciata all’uomo, e l’ *agire razionalmente*, che è ciò che l’intelligenza artificiale riesce a fare.

Accanto allo sviluppo del *machine learning*, continuano gli studi sulle tecniche di rappresentazione della conoscenza, il ragionamento automatico, la pianificazione e il supporto decisionale, con le quali si sviluppano sistemi trasparenti, in grado di spiegare le proprie decisioni in un linguaggio comprensibile a un esperto umano<sup>40</sup>.

Gli aspetti sinora analizzati concernono l’IA come disciplina scientifica, cioè come tecnica usata per la produzione di sistemi di IA. Tale disciplina ha portato alla creazione di agenti intelligenti o sistemi intelligenti, alcuni dei quali sono già stati citati a titolo di esempio. Va precisato che non tutti gli agenti intelligenti sono dotati di fisicità.

Infatti, nell’ambito dell’informatica, l’IA definisce le tecniche di progettazione di una macchina detta *elaboratore* che riproduce talune facoltà dell’intelligenza umana. In questo modo, la macchina si sostituisce all’uomo, ma non interagisce con l’esterno<sup>41</sup>.

Per contro, quando l’agente intelligente è dotato di fisicità, si ha un robot che agisce nel mondo fisico. Questo è l’ambito di studio della robotica cognitiva<sup>42</sup>, ove a essere progettata è una macchina detta *interelaboratore* o *robot intelligente* che, riproduce alcune facoltà dell’intelligenza umana, ma anche la capacità dell’uomo di interagire col mondo. Questa interazione si realizza per il tramite di sensori che consentono alla macchine di percepire degli *input*, cioè delle informazioni che saranno poi elaborate al fine di produrre un output attraverso i cosiddetti attuatori. La robotica cognitiva si occupa altresì della creazione di agenti software intelligenti, detti *softbot*. Questi, a differenza dei robot, agiscono nel mondo virtuale del web<sup>43</sup>.

---

le recenti dichiarazioni di un dipendente della Google sul chatbot LaMDA, N. Tiku, *The Google engineer who thinks the company’s AI has come to life*, in *Whashingtonpost.com*, 11 June 2022.

<sup>39</sup> Searle, J. R., *Is the Brain’s Mind a Computer Program?*, in *Scientific American*, January 1990, pp. 26-32.

<sup>40</sup> *Proposte per una strategia italiana per l’intelligenza artificiale*, a cura del Gruppo di esperti MISE sull’IA, Ministero dello sviluppo economico, p. 11. Il documento è disponibile online.

<sup>41</sup> M. Somalvico, F. Amigoni, V. Schiaffonati, *La grande scienza. Intelligenza artificiale*, cit., [www.treccani.it](http://www.treccani.it).

<sup>42</sup> L. Carlucci Aiello, *Intelligenza artificiale*, cit., [www.treccani.it](http://www.treccani.it).

<sup>43</sup> *Id.*, [www.treccani.it](http://www.treccani.it).

La distinzione tra robotica e intelligenza artificiale non è sempre chiara e, anzi, delle volte i due termini sono considerati equipollenti. Ad esempio, la risoluzione del Parlamento Europeo del 2017 contiene raccomandazioni rivolte alla Commissione concernenti norme di diritto civile sulla robotica. Nondimeno, a dispetto della formale intitolazione, essa pone anche dei principi riguardanti “ lo sviluppo della robotica e dell’intelligenza artificiale per uso civile”.

È certo che la robotica non si esaurisce nell’intelligenza artificiale e che quest’ultima non si limiti alla robotica<sup>44</sup>.

Ciò che accomuna l’informatica alla robotica è che l’IA si occupa in entrambi i casi dell’elaborazione di informazioni avvalendosi di algoritmi. In generale, si è soliti riferirsi all’*elaboratore* e al *robot intelligente* come *agente intelligente* o *sistema intelligente*.

Per questo, nel prosieguo adopereremo questa terminologia generica per riferirci tanto ai *software*, quanto ai *robot* intelligenti.

## **2. “Basta aggiungere dati”. Breve spiegazione del *machine learning* da parte di un giurista**

È stato già evidenziato che, tra le varie tecniche di IA, l’apprendimento automatico è quella che desta maggiore interesse in quanto consente al sistema intelligente di agire in autonomia. Di seguito tenterò di spiegare gli aspetti del *machine learning* che appaiono rilevanti per la successiva analisi giuridica della responsabilità connessa alla produzione e all’uso di sistemi di IA.

Iniziamo allora col dire che la caratteristica del *machine learning* o apprendimento automatico<sup>45</sup> sta nella capacità dell’algoritmo di creare altri algoritmi. Infatti, può accadere che non vi sia un algoritmo in grado di assolvere alcuni compiti che assegniamo a una macchina, poiché non si conosce la regola per trasformare l’*input* nell’*output* desiderato<sup>46</sup>. Questo concetto è ben comprensibile attraverso l’esempio dello *spam* nella casella di posta elettronica. Classificare una e-mail come spam non è un dato sempre uguale, invariabile, poiché ciò che è spam per un soggetto potrebbe non esserlo per un altro. Pertanto, non disponiamo di una conoscenza che ci consente di definire con certezza

---

<sup>44</sup> M. Bassini, L. Liguori, O. Pollicino, *Sistemi di intelligenza artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, in *Intelligenza artificiale, protezione dei dati personali e regolazione*, a cura di F. Pizzetti, 2018.

<sup>45</sup> K. P. Murphy, *Machine learning. A probabilistic perspective*, The MIT Press, Massachusetts, 2012.

<sup>46</sup> E. Alpaydin, *Introduction to Machine Learning*, The MIT Press, Cambridge, 2010, p. 1.

cosa *a priori* rientri nel concetto di ‘spam’. Tuttavia, la conoscenza può essere ricavata dai dati. A questo punto interviene il *machine learning*. L’addestratore dell’algoritmo classifica una serie di e-mail come spam e sarà compito del sistema di IA estrarre autonomamente un algoritmo per adempiere il compito assegnatogli.

A questa attività creativa l’algoritmo giunge senza un vero “ragionamento”, ma attraverso l’elaborazione di dati. Ad esempio, un sistema di traduzione non comprende il significato della frase da tradurre, eppure riesce a tradurla perché, dall’analisi dei testi inseriti nel sistema, esso comprende che un numero elevato di volte a una data parola ne segue un’altra. Per questo è stato detto che le decisioni algoritmiche sono “procedurali” e non “sostanziali”, atteso che non vi è una comprensione simbolica o semantica della decisione assunta<sup>47</sup>.

Un paragone che aiuta a spiegare il *machine learning* è quello dell’agricoltura<sup>48</sup>. L’agricoltore pianta i semi, concima il terreno e attende che il seme germogli. Nell’apprendimento automatico, l’agricoltore è il *learner*, esperto di *machine learning*, che utilizza degli algoritmi di apprendimento, equiparabili ai semi, che si “nutrono” di dati, come un seme si nutre delle sostanze che trova nel terreno. Da ciò deriverà il programma appreso, equiparabile, a sua volta, alla pianta germogliata.

Il punto di forza di questa tecnica, pertanto, sono i dati: è attraverso di essi che il sistema riesce a imparare.

Facendo il paragone con un bene di consumo, se un sistema che funziona attraverso la tecnica del *machine learning* si trovasse esposto sugli scaffali di un supermercato, nell’etichetta ci sarebbe scritto “basta aggiungere dati”<sup>49</sup>.

Il *machine learning*, pertanto, consiste in un algoritmo che è in grado di apprendere partendo dai dati e riesce a affinare le proprie conoscenze attraverso l’esperienza. A sua volta, un algoritmo è una sequenza di istruzioni che deve essere seguita per trasformare un *input* in un *output*<sup>50</sup>.

Grazie al *machine learning* si tenta di imparare dal passato qualcosa che possa valere anche per il futuro<sup>51</sup>. L’apprendimento è finalizzato alla predizione, alla risoluzione di casi diversi da quelli analizzati, che si potranno verificare in futuro. Questo processo è

---

<sup>47</sup> D. Cardon, *Le Pouvoir des algorithmes*, in *Pouvoirs*, 164, 2018, p. 67.

<sup>48</sup> P. Domingos, *L’algoritmo definitivo*, Torino, Bollati Boringhieri, 2016, p. 29.

<sup>49</sup> Anche questo paragone è di P. Domingos, *L’algoritmo*, cit., p. 29

<sup>50</sup> E. Alpaydin, *Introduction to Machine Learning*, The MIT Press, Cambridge, 2010, p. 1.

<sup>51</sup> P. Domingos, *L’algoritmo definitivo*, cit., p. 86.

definito *generalizzazione*: dato un insieme di informazioni iniziali, bisogna estrapolare una regola che sia idonea a prevedere e risolvere casi futuri non ancora analizzati<sup>52</sup>: il *machine learning* mira a prevedere un certo risultato.

Secondo una nota definizione, “*A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with the experience E*”<sup>53</sup>.

L'apprendimento è così definito alla luce di tre parametri: il compito o obiettivo assegnato al sistema, il livello di successo raggiunto nella prestazione e l'esperienza. Su ognuno di essi ci soffermeremo di seguito.

## 2.1 *Tasks*: quali compiti possono essere affidati a una macchina ‘che apprende’

Procedendo con ordine, va anzitutto evidenziato che nell'ambito del *machine learning*, a ogni algoritmo viene assegnato un compito (*task*), che viene adempiuto attraverso il raggiungimento di un *output*. Quest'ultimo è il risultato dell'elaborazione dei dati processati dall'algoritmo, i quali costituiscono l'*input*.

I compiti affidati a un algoritmo di apprendimento automatico possono essere diversi<sup>54</sup>. All'algoritmo può essere demandato il compito di classificare un certo *input* entro una certa categoria, ad esempio classificare la foto di un cane husky nella categoria ‘cane’ o ‘lupo’ (*classification*). Si tratta del riconoscimento delle immagini, tecnologia usata, ad esempio, per il riconoscimento facciale e per il riconoscimento degli oggetti nelle auto a guida autonoma<sup>55</sup>.

Vi sono poi mansioni predittive, la cosiddetta *regression*, per effetto della quale all'algoritmo è domandato di prevedere un valore numerico a fronte di un dato *input*. Si

---

<sup>52</sup> Riportano questa definizione in linguaggio matematico M. Belkin, D. Hsu, S. Ma, S. Mandal, *Reconciling modern machine learning practice and the classical bias-variance trade off*, in *PNAS*, vol. 116, n. 32, 2019, p. 15849-14854.

<sup>53</sup> T. Mitchell, *Machine Learning*, McGraw Hill, 1997, p. 2

<sup>54</sup> I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*, MIT Press, 2016, p. 98 ss. Disponibile nella versione ebook <http://www.deeplearningbook.org>. Gli autori forniscono una elencazione delle attività riservate a un algoritmo descrivendo anche le funzioni matematiche che stanno alla base.

<sup>55</sup> L'impiego di questa tecnologia è veramente vario. Di recente, una ricerca dell'Università di Copenhagen condotta con l'impiego della tecnologia del riconoscimento di immagini, ha dimostrato che nel deserto del Sahara occidentale e nel Sahel si trovano 1,8 miliardi di alberi. Questi sono stati individuate attraverso l'analisi di immagini satellitari da parte di un software che impiega la tecnologia di *deep learning*, addestrato a riconoscere gli alberi. Si tratta di una notizia accolta con grande interesse dal mondo scientifico. M. Hornbek, *AI has just revealed there are over 1.8 billion trees in the Sahara Desert* in <https://www.weforum.org/agenda/2020/10/ai-billion-trees-sahara-desert-technology/>.

pensi, in quest'ultimo caso, alla previsione della richiesta risarcitoria che avanzerà la vittima che ha subito un danno. A differenza della *classification*, nella *regression* il risultato è un numero<sup>56</sup>, un valore che non è predeterminato tra due possibili categorie (tornando all'esempio, categoria 'cane' o 'lupo', oppure 0 o 1), ma viene determinato "liberamente" dall'algorithm, non essendo inquadrabile entro delle categorie prestabilite.

La previsione di un dato futuro caratterizza anche i *learning associations tasks*, ove l'obiettivo è trovare una regola che associ due soggetti. È il caso degli acquisti da parte dei consumatori, allorché l'azienda sia interessata a capire quale prodotto sarà scelto da un dato consumatore in futuro. In base agli acquisti pregressi è chiesto all'algorithm di prevedere con quale probabilità un dato consumatore acquisterà un certo prodotto. Il quesito si esprime simbolicamente in questi termini:  $P(Y|X)$ , dove Y è il nuovo prodotto che dovrebbe essere acquistato dal consumatore che ha già acquistato il prodotto X. L'algorithm troverà un'associazione in termini probabilistici, ad esempio dirà:  $P(Y|X) = 0.7$ . Facendo un esempio concreto, se Y corrisponde a 'chips' e X a 'birra', allora il 70% dei consumatori che ha acquistato la birra acquisterà anche le chips<sup>57</sup>. È questo il modo in cui funziona l'algorithm di Amazon, che associa ai prodotti acquistati ulteriori prodotti suggeriti alla luce delle ricerche già effettuate dall'utente.

A livello linguistico, i compiti possono essere la trascrizione di dati in un certo linguaggio, ovvero la traduzione di un testo da una lingua a un'altra.

Tra le altre funzioni che possono essere assegnate a un algorithm vi sono la rilevazione di anomalie in un sistema, compito che consente di sventare i tentativi di frode nell'uso di carte di credito.

Come già precisato, il risultato o *task* è raggiunto dall'algorithm attraverso il processo di apprendimento. Quest'ultimo, dunque, è lo strumento col quale il fine viene raggiunto e non il fine medesimo.

La capacità dell'algorithm di imparare una *regola* a partire dai dati è indicata come *knowledge extraction*<sup>58</sup>. La regola è un *modello* che spiega i dati analizzati dal sistema di IA. Tornando all'esempio del consumatore, l'algorithm imparerà – da solo- che chi sceglie il prodotto X sceglierà probabilmente anche il prodotto Y.

---

<sup>56</sup> E. Alpaydin, *Introduction to Machine Learning*, cit., p. 9.

<sup>57</sup> E. Alpaydin, *Introduction to Machine Learning*, cit., p. 4.

<sup>58</sup> E. Alpaydin, *Introduction to Machine Learning*, cit., p. 8.

## 2.2 Performance measure: la valutazione delle prestazioni dell'algoritmo

La qualità del processo di apprendimento, ovvero della prestazione del sistema di IA, è valutata attraverso il parametro  $P$ , che sta per *performance measure*.

In generale, si identifica questa attività con l'espressione “*machine learning model evaluation*”<sup>59</sup>. Oggetto della valutazione è un *modello di apprendimento*, cioè uno dei modi in cui l'algoritmo o il sistema di IA può essere strutturato. Esistono, infatti, diversi modelli o tecniche di apprendimento, che esamineremo in seguito.

L'obiettivo di un produttore di sistemi di IA è individuare il modello che offre la prestazione migliore, dato un certo compito  $T$ . Per fare ciò, la scelta del modello è preceduta dalla valutazione di più modelli messi a confronto tra loro<sup>60</sup>.

La valutazione del modello è fatta in momenti diversi. In primo luogo, essa è effettuata nella fase di addestramento, e, in secondo luogo, dopo che il modello è stato addestrato.

Per fare ciò, l'insieme di dati (*dataset*) di cui dispone il *learner* viene suddiviso in tre sottoinsiemi, ognuno con dati diversi. Il primo insieme è definito *training dataset* e raccoglie i dati utilizzati per l'addestramento. Il secondo insieme è detto *validation dataset*. Esso comprende dati diversi da quelli impiegati nella fase di addestramento, i quali vengono utilizzati per testare il comportamento della macchina dinanzi a dati nuovi<sup>61</sup>. Infine, dopo che l'addestramento è completato, l'accertamento è ripetuto un'ultima volta su un altro insieme di dati, il *test dataset*.

Il modello viene quindi valutato ben due volte: durante l'addestramento, viene testato sul *set di validazione*, successivamente, una volta che il modello è stato addestrato, si procede al test finale sui dati raggruppati nel *set di test*.

Vi sono vari parametri per valutare la prestazione, di cui tener conto a seconda del compito ( $T$ ) assegnato al sistema. Ciò significa che, a seconda del *task*, si usano parametri diversi, cioè una differente *performance metric*.

---

<sup>59</sup> Per approfondimenti di tipo tecnico si rinvia a A. Zheng, *Evaluating Machine Learning Models*, O'REILLY, 2015.

<sup>60</sup> P. Mehta, M. Bukov, C.-H. Wang et al., *A high-bias, low-variance introduction to Machine Learning for physicists*, in *Physics Reports 810* (2019) 1–124, p. 7.

<sup>61</sup> R. Marmo, *Algoritmi per l'intelligenza artificiale*, Hoepli, Milano, 2020, p. 151.

## 2.2.1 La valutazione dell'algoritmo che procede alla classificazione

Posto che i parametri di valutazione cambiano a seconda del compito  $T$  che il modello è chiamato a svolgere, di seguito ci soffermeremo sui parametri cui si fa ricorso per valutare la prestazione di un algoritmo impiegato per la *classificazione*.

### 2.2.1.1. Accuratezza

Per la classificazione, si fa ricorso a *performance metrics* come l'*accuracy* o la *long-loss*<sup>62</sup>. Nella classificazione binaria, l'*output* fornisce solo due numeri da associare a due classi, mentre nella classificazione multi-classe vi sono tanti numeri di *output* quante sono le classi<sup>63</sup>. Ebbene, l'*accuracy* si riferisce alla percentuale con la quale il sistema fornisce il risultato corretto<sup>64</sup>.

Nella fase di addestramento, l'accuratezza consiste nel rapporto tra il numero delle risposte corrette date dalla macchina e il numero totale dei dati che ha a disposizione nel *dataset*.

$$\text{accuracy} = \text{correct predictions} / \text{total data points}$$

Per agevolare la comprensione, prendiamo in considerazione il caso della classificazione dei sospetti malati di tumore. Si tratta di una classificazione binaria, poiché alla macchina chiediamo di classificare il dato inserito, riguardante il singolo paziente, come negativo, in caso di paziente sano, o positivo, in caso di paziente in cui viene riscontrata la malattia. I casi possibili di *output* ottenuti dal sistema possono essere di quattro tipi:

- a) *true positive (TP)*, laddove il sistema individui correttamente un caso positivo (si tumore);
- b) *true negative (TN)*, nel caso di corretta classificazione di un caso come negativo (no tumore);
- c) *false positive (FP)*, nel caso di errore della macchina, la quale indica un caso negativo come positivo (diagnostica erroneamente un tumore);
- b) *false negative (FN)*, quando la macchina erra scambiando un soggetto malato per uno sano (mancato riconoscimento del tumore in corso).

---

<sup>62</sup> A. Zheng, *Evaluating Machine Learning Models*, O'REILLY, 2015, p. 3.

<sup>63</sup> R. Marmo, *Algoritmi per l'intelligenza artificiale*, cit., p. 218.

<sup>64</sup> I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, cit., p. 101.

Si tratta delle risposte che possono essere ottenute in tutti i casi di classificazione binaria. Naturalmente la quantità degli esiti aumenta nel caso di classificazioni multiple. È questo il caso in cui a un sistema di guida autonoma sia richiesto di riconoscere l'ostacolo che ha davanti. L'output atteso sarà di N classi, poiché potrebbe trattarsi di un masso, un cane, un gatto, un uomo, e via enumerando. Per esigenze di semplicità espositiva, prendiamo in considerazione in questa sede il compito della classificazione binaria.

Orbene, sulla base dei possibili risultati ottenibili, la formula della *accuracy* sarà la seguente:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Maggiore sarà il valore ottenuto entro l'intervallo 0 e 1 e migliore sarà l'accuratezza<sup>65</sup>. Potrebbe ritenersi che una maggiore accuratezza sia il risultato di un insieme di dati più ampio, ma la correlazione non è automatica<sup>66</sup>. In ogni caso, la sola accuratezza si rivela un parametro insufficiente. Essa presenta dei limiti, per superare i quali si ricorre alla matrice di confusione.

### 2.2.1.2 Matrice di confusione

Il limite del parametro dell'accuratezza è non fare distinzioni tra le classi. In altri termini, non si è in grado di comprendere quanti errori di classificazione si sono verificati entro la classe 0 e quanti all'interno della classe 1. Si tratta di un dato che può essere rilevante in determinati contesti. Ad esempio, il dato in questione rileva proprio nel caso della diagnosi medica, poiché l'errore diagnostico da cui discende un falso positivo ha conseguenze ben diverse rispetto all'errore che consiste nel ritenere un soggetto sano, quando invece presenta la malattia (falso negativo).

Per tale esigenza si è soliti ricorrere alla matrice di confusione o *confusion matrix*, che consente di individuare le risposte corrette e quelle sbagliate all'interno di ogni singola classe. Ricordando il nostro esempio, la matrice di confusione sarà in grado di

---

<sup>65</sup> R. Marmo, *Algoritmi per l'intelligenza artificiale*, cit., p. 223.

<sup>66</sup> Cfr. Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, 8 aprile 2020, Appendix, lett. A, n. 5: "Although it may seem as if larger datasets provide better chances of finding recurrent patterns and correlations, accuracy rates do not automatically increase with the size of the dataset. As a result of the large number of people affected by algorithmic systems, the number of errors in the form of false positives and false negatives, and of people who are affected by these errors and inbuilt bias, will also expand, triggering additional interferences with the exercise of human rights in multiple ways".

indicare quante volte il sistema ha riconosciuto un vero positivo e quante volte ha errato in ciò, così come quante volte si è sbagliato nel classificare come positivo un negativo ovvero quante volte ha correttamente riconosciuto un negativo.

I risultati della matrice di confusione sono riportati in una tabella. Le classi reali in sede di test set 0 e 1 (cioè di quanti Vero – Falso si caratterizzava l’insieme di dati) sono indicate nelle caselle orizzontali, mentre le previsioni fatte dal sistema come 0 o 1 (Vero o Falso) nelle colonne.

n = 300	Previsti come positivi	Previsti come negativi	
<b>Reali positivi</b>	80	20	100 positivi
<b>Reali negativi</b>	5	195	200 negativi
	85	315	

**Tabella 2.1:** esempio di matrice di confusione. In alto a sinistra, il numero di casi totali, pari a 300, sottoposti al vaglio della macchina. Il numero reale, quindi effettivo, di positivi e negativi è riportato nell’ultima colonna a destra, in corrispondenza delle caselle orizzontali “Reali positivi” e “Reali negativi”. Nelle colonne centrali il numero dei casi classificati, rispettivamente, come negativi e come positivi. Pertanto, si nota che su 100 positivi, la macchina ne ha classificati correttamente solo 80, mentre 20 sono ritenuti erroneamente negativi. Per contro, su 200 casi negativi, solo 5 sono stati classificati come positivi, mentre ben 195 hanno ricevuto la corretta classificazione.

Prendiamo in considerazione il caso descritto nella tabella 2.1. Il *dataset* conteneva 100 esempi positivi (Vero) e 200 esempi negativi (falso), ma il sistema prevede solo 80 casi positivi e solo 195 negativi<sup>67</sup>. Da ciò consegue che:

- la classe dei positivi (o Vero o 0) ha un’accuratezza pari all’80%, calcolata nel modo seguente  $80/(20 + 80) = 80\%$ ;
- la classe dei negativi ha un’accuratezza pari al 97.5%, poiché  $195/(5 + 195) = 97.5\%$ .

Ciò significa che il sistema ha un’accuratezza inferiore nella predizione del risultato come positivo, mentre l’accuratezza è maggiore nel riconoscimento dei dati negativi.

---

<sup>67</sup> È questo l’esempio riportato da A. Zheng, *Evaluating Machine Learning Models*, p. 9.

Senza ricorrere alla matrice di confusione, si disporrebbe del solo dato dell'accuratezza totale, pari a 91.7%, cioè  $(80 + 195) / (100 + 200)$ .

Questi dati sono rilevanti a seconda del quesito di classificazione posto al sistema. Ad esempio, nel caso dei sistemi di diagnostica tumorale, è importante sapere che la percentuale di errore nella previsione dei casi negativi è più bassa rispetto a quella dei falsi positivi: un falso positivo potrà essere sventato nel corso dei successivi accertamenti, mentre un falso negativo è più difficile da individuare, poiché è probabile che il paziente non si sottoporrà a ulteriori accertamenti confidando nel fatto di godere di buona salute.

L'accuratezza è un valore di particolare importanza se si pensa al rapporto uomo-macchina.

Anzitutto, se in fase di test si rileva che la macchina ha una accuratezza troppo bassa, potrebbe concludersi che il prodotto non è idoneo per essere messo in commercio. Collocandoci dal punto di vista della responsabilità dell'utente consumatore, va osservato che se la macchina è in grado di fornire una risposta circa il livello di accuratezza della soluzione da essa offerta, l'uomo sarà in grado di valutare con maggiore consapevolezza se seguire o meno il responso dato dal sistema di IA. Si pensi al sistema di IA progettato per riconoscere una certa patologia indicato nell'esempio. A seconda del livello di accuratezza del risultato ottenuto, il medico ben potrà decidere se effettuare ulteriori accertamenti in quella direzione o se discostarsi dalla consulenza data dalla macchina.

Oltre all'accuratezza, può essere misurato il livello di errore (*error rate*), cioè la percentuale che indica la probabilità che la risposta fornita dalla macchina sia errata. L'argomento è affrontato nel paragrafo sui temi dell'*underfitting* e dell'*overfitting*, al quale si rinvia.

### **2.2.1.3 Oltre l'accuratezza: precisione, sensibilità, specificità, *fall-out***

I dati di cui disponiamo una volta ricostruita la matrice di confusione, ci consentono di disporre di risultati attinenti a metriche diverse dalla *accuracy*, utili per individuare parametri ulteriori. Si tratta di parametri che hanno riguardo ai valori predetti dal sistema di IA. Considerando la tabella sopra riportata, la lettura è effettuata, a seconda del parametro di riferimento, in verticale ovvero in orizzontale, come mostrato nella tabella 2.2.

Iniziamo allora dalla *precisione* (*precision*). Essa indica il rapporto tra il numero dei casi che il sistema ha correttamente individuato come positivi e il valore totale dei casi indicati come positivi dal sistema, inclusi anche i falsi positivi.

$$\textit{Precisione} = TP / (TP + FP)$$

Si tratta di un valore che attiene a una sola classe, cioè quella dei positivi (o vero) e relativa alla sola colonna della previsione fatta dal modello, poiché non rileva il numero reale dei positivi. Pertanto, i dati vanno estratti dalla prima colonna della parte della tabella relativa alla previsione della macchina.

Nel caso della precisione, poiché l'intervallo è compreso tra 0 e 1, si osserva che tanto più grande è il valore, quanto maggiore sarà la precisione del sistema, poiché significa che il valore *FP* (false positive) è minore. Difatti, se  $FP = 0$ , allora il rapporto  $TP / (TP + FP)$  sarà uguale a 1. Per converso, tanto più grande è il valore di casi erroneamente classificati come positivi, tanto più basso sarà il livello di precisione.

Per ritornare all'esempio sopra citato, il sistema aveva correttamente individuato come positivi soltanto 80 casi su 100; ha inoltre indicato 5 casi negativi come positivi. Di conseguenza, la precisione sarà calcolabile nel modo seguente:  $80 / (80+5) = 0,9$ . Si nota che il livello di precisione è abbastanza alto perché prossimo al valore 1. In questo caso, pertanto, è basso il rischio di errore, poiché il valore *FP* contiene pochi casi e si riduce la probabilità che il risultato dato dal sistema nel singolo caso sia uno di essi.

Diversa dalla precisione è la *sensibilità (recall)*. Essa indica la *percentuale di veri positivi correttamente classificati dal sistema*, risultante dal rapporto tra il numero di casi correttamente individuati come positivi e il dato reale dei casi positivi. Al denominatore, pertanto, configurerà la somma dei veri positivi e dei falsi positivi:

$$\textit{Sensibilità} = TP / (TP + FN)$$

Dato l'intervallo compreso tra 0 e 1, una maggiore sensibilità denoterà una migliore accuratezza del sistema. Anche in questo caso, se il valore dei falsi negativi è piccolo, allora la classificazione dei dati entro i veri positivi si avvicina al valore 1, coincidendo con essa laddove  $FN = 0$ . Se  $FN$  è un valore alto, cresce il rischio di errore, poiché molti casi realmente positivi non sono stati individuati dal sistema.

Nell'esempio sopra riportato, la sensibilità sarà calcolata nel seguente modo:  $80 / (80 + 20) = 0,8$ . I dati vanno estrapolati leggendo la prima linea orizzontale della tabella 2.2.

Il parametro speculare alla sensibilità, ma riferito alla *percentuale dei casi correttamente classificati come negativi dal sistema*, è la *specificità*.

$$\text{Specificità} = TN / (TN + FN)$$

Con riguardo all'esempio da noi considerato, la specificità sarà calcolata leggendo la seconda riga orizzontale, pertanto:  $195 / (195 + 5) = 0,97$ . Approssimandosi al valore 1, aumenta l'accuratezza, poiché significa che è molto basso il valore FN di casi positivi erroneamente qualificati come negativi.

Infine, è altresì possibile calcolare la percentuale di "falsi allarmi", cioè di casi negativi erroneamente qualificati come positivi dall'agente intelligente. Questo valore è denominato *fall-out*.

$$\text{Fall-out} = FP / (FP + TN)$$

Una migliore accuratezza corrisponderà a un valore di *fall-out* basso, poiché ciò significa che la percentuale di errore è molto bassa in quanto elevato è il numero di corretti casi negativi indicato al denominatore, mentre è basso il valore di falsi positivi che sta al numeratore.

Tornando al nostro esempio:  $20 / (20 + 195) = 0,09$ .

n = 300	Previsti come positivi	Previsti come negativi	Sensibilità
	80	20	
Reali positivi	80	20	100 positivi
Reali negativi	5	195	200 Negativi
	85	215	
	Precisione	Fall-out	
	Specificità		

**Tabella 2.2** Gli insiemi indicano i dati necessari per calcolare i rispettivi valori.

Non sempre tutti questi parametri sono rilevanti. Talvolta soltanto alcuni o uno di essi verranno presi in maggiore considerazione rispetto agli altri. Ciò dipende dal *task*, cioè dal compito affidato alla macchina.

Ad esempio, nel caso di compiti di diagnostica di malattie mortali o di virus trasmissibili (si pensi alla diagnostica dell'ebola, dell'HIV, del Covid-19), ciò che più rileva è avere un numero di falsi negativi basso. Pertanto, il parametro su cui porre maggiore attenzione è la sensibilità, che deve essere prossima a 1 per evitare un numero elevato di falsi negativi, così riducendo il rischio di non informare della patologia un soggetto malato.

Laddove la classificazione miri a individuare i video adatti ai minori, bloccando quelli considerati pericolosi, deve prediligersi un modello che dia la garanzia di zero falsi positivi, da intendersi come positivi quelli adatti ai minori e negativi quelli non adatti. Quindi serve un livello elevato di precisione, mentre può essere tollerato un livello basso di sensibilità, cioè di percentuale di veri positivi, poiché a poco rileva che il sistema, sbagliando, blocchi video sicuri (veri positivi) classificandoli come negativi (sono i falsi negativi)<sup>68</sup>.

I dati risultanti dalla matrice di confusione possono essere riassunti in un grafico dove viene tracciata la *curva ROC (Receiver Operating Characteristic Curve)*. La curva ROC consente di scegliere la migliore soglia di attivazione per una data decisione.

Queste valutazioni consentono di modulare il livello di rischio consentito al produttore. Non ogni parametro avrà lo stesso livello di rischio consentito per ogni tipo di prodotto. Ciò inciderà sulla responsabilità penale del produttore, assumendo rilevanza già in sede di accertamento della causalità e della responsabilità colposa.

Per questo motivo, il produttore dovrà occuparsi di svolgere un'attenta analisi preliminare al fine di definire gli obiettivi da raggiungere e stabilire i parametri cui dare maggiore importanza e la soglia entro la quale è ammesso l'errore del sistema.

---

<sup>68</sup> R. Marmo, *Algoritmi per l'intelligenza artificiale*, cit., p. 225.

Le tecniche di valutazione qui illustrate si riferiscono ai casi di *classification*. A seconda del compito assegnato, varia la tecnica di valutazione del modello. Si pensi ai compiti di *regression*, per i quali non si fa riferimento all'*accuracy*, poiché la prestazione viene solitamente valutata ricorrendo al *root-mean-square error (RMSE)* o *root-mean-square deviation*<sup>69</sup>. Ben si comprende come l'argomento relativo alla valutazione dei modelli di *machine learning* sia molto vasto. Per questa ragione, qui abbiamo deciso di soffermarci brevemente solo su alcuni aspetti della valutazione degli algoritmi di classificazione. Per gli altri sistemi di valutazione degli ulteriori modelli, si rinvia alla letteratura sull'argomento<sup>70</sup>.

### 2.2.2 Sbagliando si impara. *Bias* e *varianza*

Come già accennato, un altro parametro da prendere in considerazione durante la fase del training e quella successiva della valutazione del modello è l'errore.

Questo va calcolato già nella fase di addestramento, in rapporto al *training dataset*, cioè all'insieme di dati utilizzati per l'addestramento. Successivamente, il valore ottenuto va confrontato con il livello di errore fatto dal sistema in rapporto al *validation dataset*. Tale test concerne dati diversi da quelli impiegati nella fase di addestramento, dunque sconosciuti al sistema di IA. L'utilità di questo test sta nel verificare il comportamento della macchina dinanzi a dati nuovi. Infine, l'accertamento è ripetuto su un altro insieme di dati utilizzati per il test finale, il *test dataset*. Anche in questo caso, i dati sono sconosciuti al modello durante la fase di addestramento.

Se l'accuratezza del modello sul *set* di test non è soddisfacente, allora bisogna capire dove è collocato l'errore.

L'errore può essere di due tipi: sistematico o variabile sull'insieme di dati. Nel primo caso ci si riferisce all'errore come *bias* o distorsione, mentre nel secondo caso si impiega il concetto di *varianza*<sup>71</sup>.

---

<sup>69</sup> A. Zheng, *Evaluating Machine Learning Models*, cit. p. 15, cui si rinvia per ulteriori approfondimenti.

<sup>70</sup> Una visione di insieme è data da R. Marmo, *Algoritmi per l'intelligenza artificiale*, cit., nel capitolo 10, *Valutazione del modello*, p. 231-236, con riferimento al *clustering* e alla *regression*.

<sup>71</sup> Il termine "varianza" fu utilizzato per la prima volta dallo statista inglese Ronald Fisher (1890-1962), in occasione di un lavoro di analisi delle colture sviluppate negli anni '40 dell'800. Inventò, allora, l'analisi della varianza, misura di dispersione per rappresentare la variabilità dei dati rispetto alla media, v. A. Ruiz de Villa Robert e E. Puertas Prats, *I big data, il machine learning e la statistica*, volume II della collana *La matematica che trasforma il mondo*, RBA editore, 2020, p. 29.

Se il modello continua a commettere sempre lo stesso errore, vi è un problema di *bias*; all'opposto, se l'errore c'è, ma non è regolare, il modello ha un problema di varianza<sup>72</sup>. Così diremo che un orologio che è sempre indietro di un'ora ha un *bias* alto, ma una varianza bassa; se invece l'orologio alterna in modo casuale anticipo e ritardo, ma in media segna l'ora corretta, il *bias* sarà basso, mentre avrà una varianza alta<sup>73</sup>.

Chi si occupa dell'addestramento dell'algoritmo dovrà analizzare l'errore, correggerlo e testare nuovamente l'algoritmo, per il quale, trattandosi di un sistema di apprendimento, vale la regola 'sbagliando si impara'.

L'analisi dei concetti, speculari, di accuratezza e errore spiega che l'addestratore dell'algoritmo ha la possibilità di valutare la prestazione dello stesso e di individuare una soglia entro la quale l'errore è ritenuto tollerabile. Ampliando l'insieme dei dati di *input*, l'errore sarà sempre più probabile e l'accuratezza sempre inferiore. È difficile che un sistema complesso, che si confronta con un'elevata mole di dati, abbia un errore pari a zero<sup>74</sup>. Tuttavia, è possibile individuare la soglia oltre la quale l'errore non è più tollerabile. Ciò significa che oltre quella soglia l'algoritmo non è ancora pronto per essere messo in commercio e che dovrà essere sottoposto a ulteriori fasi di addestramento.

### **2.2.3 Underfitting e overfitting. Come cucire l'abito perfetto**

Nel calcolo dell'errore si possono verificare tre scenari: *underfitting*, *overfitting*, *appropriate capacity*. Procediamo con ordine.

Durante la fase di addestramento sui dati del *training set*, uno dei compiti dell'addestratore è ridurre il livello di errore. Quindi attraverso più test si verifica se la macchina ha imparato a meglio riconoscere i dati di *input* dando un risultato più preciso, con un livello di errore di volta in volta meno elevato. Se il sistema non è in grado di migliorare nella fase di addestramento, il *training error* continuerà ad avere un valore molto alto. Questo è un caso di *underfitting* (o sotto-adattamento), così detto poiché il

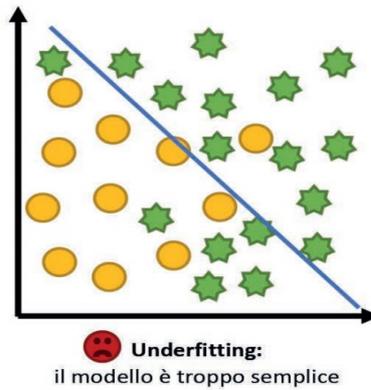
---

<sup>72</sup> Per la definizione matematica e statistica dei concetti di *bias* e varianza si rinvia a T. Mitchell, *Machine learning*, cit., p. 137 ss.

<sup>73</sup> L'esempio è tratto da P. Domingos, *L'algoritmo definitivo*, cit., p. 106.

<sup>74</sup> Questo, del resto, è il limite dei sistemi su base statistica, come si legge nella *Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems* dell'8 aprile 2020: "While it is often argued that the costs are offset by gains in rationalisation and accuracy, it is important to note that most algorithmic systems are based on statistical models in which errors form an inevitable part, sometimes with feedback loops that maintain, replicate and reinforce pre-existing biases, errors and assumptions", p. 5.

sistema non riesce a adattarsi ai dati che ha a disposizione, come se si trattasse di un abito cucito sempre male.



**Grafico 2.3.** Esempio di *underfitting*: in un caso di classificazione, il modello non è in grado di distinguere i cerchietti gialli dalle stelline verdi. Si riscontra un problema di *underfitting*.

L’obiettivo del *trainer*, pertanto, sarà tentare di cucire l’abito su misura, riducendo il livello di errore e, quindi, eliminando l’*underfitting*.

Tuttavia, una maggiore esattezza con riguardo ai dati del *training set* non assicura un sistema più performante. Come già osservato, la valutazione della macchina va fatta su dati diversi da quelli utilizzati nel *training set*, al fine di comprenderne la capacità di generalizzazione, cioè di verificare come il sistema si comporta con dati nuovi, cioè di appurare quanto bene i concetti appresi da un modello di *machine learning* si applicano a esempi diversi da quelli analizzati durante la fase di apprendimento.

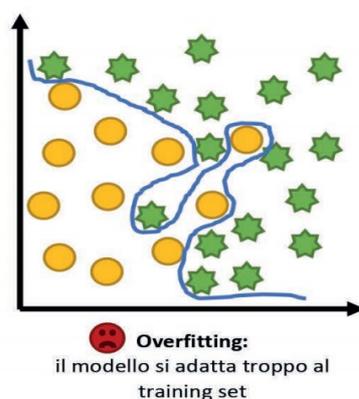
Pertanto, un livello di errore molto basso rapportato ai dati usati durante la fase di addestramento non assicura un livello altrettanto basso di errore sui dati utilizzati per il *test* finale. Difatti, può accadere che il sistema impari a classificare alla perfezione i dati già conosciuti, poiché ne conosce ogni singola caratteristica, ma poi non riesca ad avere una stessa capacità di analisi con i dati del *test* finale, dove il livello di errore sarà quindi più elevato. Il sistema ha quindi generalizzato in maniera eccessiva i soli dati che conosce. Questo è un problema di *overfitting*, caratterizzato da un’ampia divergenza tra l’errore verificatosi durante l’addestramento (*training error*) e l’errore riscontrato durante il *test* conclusivo (*test error*)<sup>75</sup>. In altri termini, se il modello raggiunge un livello di errore in

---

<sup>75</sup> Nota P. Domingos che “neanche gli esseri umani sono immuni all’*overfitting*”. Riporta a tal proposito l’esempio della bambina con pelle bianca che, incontrando al centro commerciale una bambina ispanica, si rivolge alla madre dicendo “guarda, mamma, una bebè cameriera!”. In questo caso, ciò che appare come

fase di addestramento pari a zero, allora non sarà in grado di generalizzare con dati nuovi<sup>76</sup>.

Per tornare alla metafora sartoriale, se il sarto cuce alla perfezione l'abito su un manichino, è altamente probabile che l'abito non possa essere indossato da una persona reale, che ha ragionevolmente misure diverse rispetto a quelle del fantoccio dalle sembianze umane inverosimili. Il sarto ha generalizzato male: focalizzandosi esclusivamente sul manichino, ha pensato che quelle misure potessero andar bene anche per un essere umano.



**Grafico 2.4.** Esempio di *overfitting*: in un caso di classificazione, il modello impara perfettamente a classificare i dati del *training set*, ma ciò comporta una minore capacità di generalizzazione, con un ampio margine di errore nel successivo *validation test* o test finale su dati non analizzati prima.

Il rischio di *underfitting* ovvero di *overfitting* è verificabile attraverso il parametro della *capacity*. Per capacità di un modello (*model capacity*) si intende l'abilità dello stesso di fornire la corretta risposta in rapporto ai dati analizzati<sup>77</sup>.

Come si è già accennato nel trattare il parametro della accuratezza, la misura dell'errore è un parametro specularmente opposto all'*accuracy*: il primo misura il dato negativo (il livello di errore), l'altra il dato positivo (quanto l' algoritmo riesce nel suo compito). Ciò significa che in un caso di *overfitting*, il livello di *accuracy* sarà più elevato in sede di *test set*, mentre sarà discendente in sede di *validation test*<sup>78</sup> o del *test* finale,

---

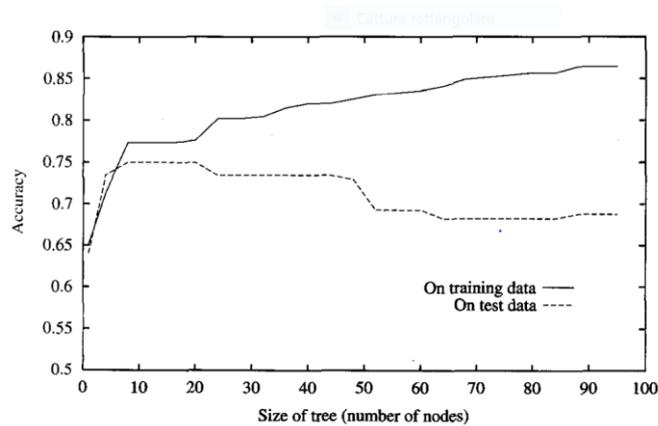
razzismo innato è, invero, un problema di *overfitting*, poiché la bambina ha generalizzato in modo eccessivo dalle poche cameriere ispaniche che ha visto nella sua vita. In *L'algoritmo definitivo*, cit., p. 99.

<sup>76</sup> M. Belkin, D. Hsu, S. Ma, S. Mandal, *Reconciling modern machine learning practice and the classical bias-variance trade off*, in *PNAS*, vol. 116, n. 32, 2019, p. 15849-14854.

<sup>77</sup> I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, cit., p. 110.

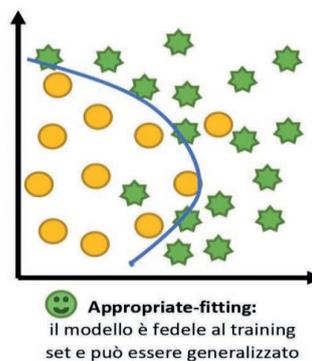
<sup>78</sup> T. Mitchell, *Machine learning*, cit. p. 66 ss. Dove si affronta il problema dell'*overfitting* in un sistema di alberi di decisione.

come si vede nel grafico 2.5 sottostante. In altri termini, all'aumentare dell'*overfitting*, corrisponde una diminuzione dell'*accuracy*.



**Grafico 2.5.** Esempio di *overfitting* in un modello di albero di decisione proposto da T. Mitchell nella sua monografia '*Machine learning*'. Sull'asse delle ascisse sono riportati il numero di nodi di un albero di decisione, che aumentano all'aumentare dei dati sottoposti al vaglio dell'algoritmo. Si nota che, all'aumentare dei nodi (quindi dei dati) l'accuratezza del modello misurata sugli esempi impiegati nella fase di addestramento aumenta. Tuttavia, a fronte di ciò, non vi è un pari incremento sugli esempi estranei all'insieme di addestramento, per i quali si nota che l'accuratezza dopo una fase ascendente, conosce un andamento calante.

L'obiettivo dell'addestratore dell'algoritmo è far in modo che esso raggiunga un livello di capacità appropriata (*appropriate capacity*), cioè che riesca nell'operazione di generalizzazione, applicando i concetti appresi durante l'addestramento a dati nuovi.



**Grafico 2.6** Il modello ideale dovrebbe evitare tanto l'*underfitting* quanto l'*overfitting*, come mostrato nel grafico.

In conclusione, vi è un margine di errore tollerabile, espressione di un bilanciamento tra il problema dell'*underfitting* e quello opposto dell'*overfitting* (ottima capacità di classificazione nel *test set*, ma scarsa capacità di generalizzazione).

Il quesito che va posto, allora, è a chi spetta stabilire il margine di tollerabilità dell'errore. Non è forse eccessivo lasciare una totale discrezionalità al produttore nel fissare la soglia entro la quale l'errore può essere ammesso?

### **2.3 *Experience*: in che modo può apprendere un algoritmo**

Terzo e ultimo parametro adoperato nella definizione da noi utilizzata è l'*esperienza* (*E*).

L'algoritmo acquisisce una sua esperienza in due momenti diversi. Anzitutto, vi è la fase di addestramento o *training*. L'algoritmo ha a disposizione un insieme di dati (*dataset*) di cui si "nutre" – per rievocare la metafora dell'agricoltura-. Successivamente, l'algoritmo è testato su un insieme di dati diverso, per verificare se ha raggiunto un livello di accuratezza generalizzato su dati non ancora analizzati<sup>79</sup> (*data test*).

Come più volte evidenziato, è importante che il sistema sia testato su un *set* di dati diverso rispetto a quello impiegato durante l'addestramento, poiché il produttore deve riuscire a prevedere come il sistema si comporterà in presenza dei dati che saranno acquisiti durante il suo utilizzo da parte del futuro consumatore.

Questa è la fase dell'apprendimento, che è definito automatico se l'agente intelligente riesce a acquisire una sua esperienza senza alcun comando umano.

L'apprendimento può essere di due tipi: supervisionato o non supervisionato.

#### **2.3.1 *Supervised vs unsupervised learning*. L'apprendimento per rinforzo**

Se consideriamo il compito della classificazione, i dati di cui si compone il *dataset* possono essere già classificati dal *learner*, cioè dall'addestratore dell'algoritmo.

Pensando al compito di classificare una immagine entro la categoria 'lupo' o 'cane husky', all'algoritmo saranno sottoposte una serie di foto già classificate come foto che ritraggono un lupo o un cane. Ciò significa che ogni elemento dell'insieme di dati è già previamente catalogato e l'algoritmo deve limitarsi a imparare le caratteristiche di ciascuna categoria.

Questo tipo di apprendimento è definito supervisionato, proprio perché l'operazione di classificazione è già svolta dal *learner*, il quale "etichetta" ogni singola immagine.

---

<sup>79</sup> E. Alpaydin, *Introduction to Machine Learning*, cit., p. 112.

All'algorithmo sarà richiesto di applicare questa conoscenza al successivo insieme di dati non analizzati in fase di addestramento e ricompresi nel *test set*. Il programmatore o addestratore fornisce al sistema i risultati attesi e segnala eventuali errori. Ogni volta che il sistema effettua una classificazione scorretta, la decisione dell'algorithmo è rivista alla luce dell'errore<sup>80</sup>.

La maggior parte dei sistemi di *machine learning* si caratterizza per l'apprendimento supervisionato<sup>81</sup>. Ad esempio, tanto la *classification* quanto la *regression* rientrano nel *supervised learning*<sup>82</sup>. Uno dei più diffusi algoritmi di apprendimento supervisionato sono le *Support Vector Machines* (SVM) o macchine a vettori di supporto<sup>83</sup>.

L'apprendimento supervisionato si differenzia da quello non supervisionato. In quest'ultimo caso, i dati di cui si compone il *dataset* non sono già classificati, pertanto la classificazione è interamente demandata all'algorithmo. Il programmatore non fornisce né i risultati attesi dal sistema, né messaggi di errore<sup>84</sup>. Lo scopo è trovare una regola generale (*pattern* o *cluster*) che accomuni certi *input*, cercando di capire quali caratteristiche ricorrono con frequenza: se nell'apprendimento supervisionato questo compito è affidato al programmatore, qui il sistema lo scoprirà da sé<sup>85</sup>. Questo processo di individuazione di pattern in statistica è chiamato *density estimation*<sup>86</sup>.

Tra i metodi in cui la *density estimation* trova attuazione, figura il *clustering*, il cui scopo è raggruppare i dati di *input* in gruppi (*clusters*). Ad esempio, se un'azienda dispone dei dati dei clienti che hanno acquistato i suoi prodotti, attraverso il *clustering* può riuscire a raggruppare i singoli clienti in gruppi di consumatore-tipo.

La tecnica del *clustering* è altresì utilizzata in ambito bio-informatico. È attraverso il *clustering*, ad esempio, che si individuano le sequenze di amminoacidi che compongono una data proteina<sup>87</sup>.

In una posizione mediana si colloca l'apprendimento semi-supervisionato, in cui solo alcuni dati sono già qualificati dall'addestratore.

---

<sup>80</sup> M. A. Boden, *L'intelligenza artificiale*, cit., p. 48.

<sup>81</sup> A. Zheng, *Evaluating Machine Learning Models*, cit., p. 7.

<sup>82</sup> E. Alpaydin, *Introduction to Machine Learning*, cit., p. 9.

<sup>83</sup> A. Ruiz de Villa Robert e E. Puertas Prats, *I big data, il machine learning e la statistica*, volume II della collana *La matematica che trasforma il mondo*, RBA editore, 2020, p. 88.

<sup>84</sup> M.A. Boden, *L'intelligenza artificiale*, cit., p. 49.

<sup>85</sup> M.A. Boden, *L'intelligenza artificiale*, cit., p. 49.

<sup>86</sup> E. Alpaydin, *Introduction to Machine Learning*, cit., p. 11.

<sup>87</sup> E. Alpaydin, *Introduction to Machine Learning*, cit., p. 12.

Un tipo particolare di apprendimento è l'*apprendimento per rinforzo*<sup>88</sup>. In tal caso, l'algoritmo interagisce con l'ambiente e riceve un riscontro della soluzione fornita al problema sottopostogli. Se il riscontro è positivo, l'algoritmo capisce che quella era la risposta giusta e in questo modo indirizza il suo comportamento futuro.

Diverse sono le tecniche di apprendimento elaborate nell'ambito del *machine learning*: decision tree, neural networks, logistic regression, K-nearest neighbors, support vector machine (SVM) e altri ancora. Di seguito ci soffermeremo su due di queste: gli alberi di decisione e le reti neurali. Queste ultime conoscono oggi una vasta applicazione e l'apprendimento che ne deriva prende il nome di *deep learning*.

### **3. Modelli di apprendimento. Simbolisti e connessionisti, *traditional machine learning* e *representation learning***

In apertura è già stato evidenziato che i modi in cui un sistema di IA può apprendere sono molteplici e riflettono visioni diverse dello studio della mente.

Tradizionalmente, si è assistito a una scissione tra due diversi approcci: quello dei simbolisti, da un lato, e dei connessionisti, dall'altro<sup>89</sup>. Tale dicotomia rispecchia l'evoluzione registratasi nella scienza cognitiva<sup>90</sup>.

Nella sua impostazione classica, la scienza cognitiva muoveva dalla convinzione che i processi cognitivi, ivi compresi quelli relativi al linguaggio, andassero individuati nella loro astrattezza e, dunque, dovessero essere studiati su un livello di astrazione rispetto alla loro realizzazione biologica<sup>91</sup>. Il cognitivismo, pertanto, è associato allo "studio della mente che considera i processi psicologici come l'elaborazione e il trattamento di simboli e informazioni, dove la mente viene rappresentata come un apparato capace di compiere tale elaborazione"<sup>92</sup>.

L'approccio più tradizionale della scienza cognitiva è quello dei simbolisti, secondo i quali ogni forma di intelligenza può essere rappresentata attraverso i simboli<sup>93</sup>. I simboli

---

<sup>88</sup> Russel, Norvig, *Intelligenza artificiale*, cit., cap. 20, p. 644 ss.

<sup>89</sup> A. Terrett, *Neural Networks - Towards Predictive Law Machines*, 3 Int'l j.l. & info. Tech. 94 (1995), p. 95.

<sup>90</sup> D. Marconi, *Linguaggio e scienza cognitiva*, in *Enc. Treccani, XXI Secolo, Norme e Idee*, 2009, pp. 201-210.

<sup>91</sup> *Id.*, p. 201.

<sup>92</sup> Voce *cognitivismo*, in *Treccani, Enc. della Scienza e della Tecnica*, 2008, [www.treccani.it/enciclopedia](http://www.treccani.it/enciclopedia)

<sup>93</sup> P. Domingos, *L'algoritmo definitivo*, cit., p. 117 ss.

sono, ad esempio, le regole con cui un determinato sistema apprende. Pertanto, secondo il pensiero simbolista, non rileva tanto il substrato biologico o artificiale dell'intelligenza che determina la manipolazione dei simboli, ma il simbolo stesso, cioè la regola.

Come già osservato nel par.1, l'applicazione del pensiero simbolista alla IA e, in particolare, all'attività di sviluppo dei software, ha dato luogo al filone dell'IA noto come intelligenza artificiale di tipo logico, contrapposta alla IA di tipo biologico<sup>94</sup>. In questo contesto di studio sono stati sviluppati i cosiddetti *sistemi esperti*<sup>95</sup>, la cui tecnologia è stata posta in primo piano nell'ambito della cosiddetta *ingegneria della conoscenza*<sup>96</sup>. Si tratta di programmi informatici nei quali viene riversata la conoscenza posseduta da uomini esperti in un dato settore, dopo essere stata tradotta in simboli, cioè in un linguaggio che il sistema riesce a comprendere. Questa base di conoscenza è assemblata per ogni specifica applicazione da 'ingegneri della conoscenza' con non poca fatica<sup>97</sup>, tant'è che è stata elaborata l'espressione 'collo di bottiglia' per descrivere le difficoltà che questa operazione comporta<sup>98</sup>. Posto che il programma incorpora una grande quantità di conoscenza su un problema specifico<sup>99</sup>, si è soliti definire tali sistemi come *sistemi basati sulla conoscenza o knowledge-based*.

In quest'ottica, pensando alla diagnosi medica come esempio, la conoscenza posseduta dai medici che consente loro di riconoscere l'influenza attraverso i sintomi è trasposta in regole poi trasfuse nel sistema di IA. Si assiste, pertanto, alla comunicazione e al trasferimento di conoscenza fra uomo e macchina<sup>100</sup>. La conoscenza è rappresentata attraverso la *logica*: quest'ultima è il linguaggio formale con cui i simbolisti esprimono i loro modelli. Inoltre, la conoscenza è espressa secondo un modello *top-down*, in quanto dalla regola generale si giunge alla decisione nel caso concreto, specializzando la regola fino a farla combaciare coi dati<sup>101</sup>.

---

<sup>94</sup> D. Parisi, voce *Intelligenza artificiale*, cit., in *www.treccani.it*.

<sup>95</sup> A. Terrett, *Neural Networks*, cit., p. 96.

<sup>96</sup> L'espressione "ingegneria della conoscenza" individua gli aspetti ingegneristici dell'intelligenza artificiale, che ha studiato per anni la tecnologia dei sistemi esperti. Cfr. sul punto, A. Carobene, voce *Intelligenza artificiale – Approfondimento*, in *Enc. Treccani*, cit.

<sup>97</sup> L. Carlucci Aiello, *Intelligenza artificiale*, cit., *www.treccani.it*.

<sup>98</sup> P. Domingos, *L'algoritmo definitivo*, cit., p. 89 e 90.

<sup>99</sup> Voce *Intelligenza artificiale*, *Treccani, enc. online*, *www.treccani.it*.

<sup>100</sup> R. Cordeschi, E. D'Avanzo, *Nuove prospettive nell'intelligenza artificiale*, in *Enc. Treccani, XXI Secolo, Norme e Idee*, 2009, pp. 183 - 190.

<sup>101</sup> Russel, Norvig, *Intelligenza artificiale*, cit., p. 692.

Nell'ambito del *machine learning*, l'approccio simbolista ha trovato attuazione nel modello degli alberi di decisione e in quello degli insiemi di regole. Di seguito ci soffermeremo sul primo dei due modelli.

Nonostante l'iniziale successo, il pensiero simbolista e i relativi modelli non hanno tardato a mostrare i loro limiti, riassumibili, anzitutto, nel fatto che la base di conoscenza dell'agente artificiale è limitata, poiché è costruita per uno specifico scopo<sup>102</sup> e, altresì, nella difficoltà di gestire la conoscenza basata sul senso comune<sup>103</sup>: come trasmettere a un sistema di IA il *common sense*?

Ma non solo. Altro limite della logica applicata ai sistemi di IA risiede nella necessità che le premesse da cui muove la conoscenza siano necessariamente vere: si riscontra una difficoltà nel gestire l'incertezza e l'incompletezza dell'informazione<sup>104</sup>. Un ulteriore problema si pone anche nella fase finale del ragionamento. Posto che l'esito del ragionamento logico è una risposta positiva o negativa, bisogna considerare che un concetto reale non sempre è definibile in un insieme di regole cui corrispondano valori di verità binari e, quindi, non è sempre associabile a una logica di vero-falso. Molti dati reali si inseriscono in "una vasta area grigia", poiché "il mondo non è in bianco e nero"<sup>105</sup>; pertanto, l'agente intelligente è chiamato ad affrontare situazioni non dominabili perfettamente, rinunciando alla certezza e assumendo una decisione basata sulla possibilità<sup>106</sup>.

L'impostazione simbolista ha cominciato a perdere consenso per effetto di due pressioni: quella del connessionismo e quella delle neuroscienze<sup>107</sup>.

Il termine *connessionismo* rievoca il fondamento di tale filone della scienza cognitiva, cioè l'idea per cui la conoscenza si accumuli nelle connessioni tra i neuroni, secondo il

---

<sup>102</sup> L. Carlucci Aiello, *Intelligenza artificiale*, cit., [www.treccani.it](http://www.treccani.it).

<sup>103</sup> R. Cordeschi, E. D'Avanzo, *Nuove prospettive*, cit., p. 187, A. Terrett, *Neural Networks*, cit., p. 96.

<sup>104</sup> R. Cordeschi, E. D'Avanzo, *Nuove prospettive*, cit., p. 187.

<sup>105</sup> P. Domingos, *L'algoritmo*, cit., p. 120.

<sup>106</sup> A. Carobene, voce *Intelligenza artificiale – Approfondimento*, in *Enc. Treccani online*, [www.treccani.it](http://www.treccani.it). Ma non va omissis che, per superare il problema, sono stati elaborati sistemi basati su logiche *fuzzy* (sfumate), alla base delle quali vi sono enunciati etichettati anche come *probabile – improbabile* o come *sconosciuto*, dando così respiro ai quesiti ingabbiati nella logica binaria *vero – falso* M. A. Boden, *L'intelligenza artificiale*, Bologna, Il Mulino, 2018, p. 40.

<sup>107</sup> D. Marconi, *Linguaggio e scienza cognitiva*, cit., p. 201.

principio dell'apprendimento associativo. Si tratta della cosiddetta *regola di Hebb*, dal nome dello psicologo canadese il cui pensiero è alla base del connessionismo<sup>108</sup>.

I connessionisti contestano la tesi simbolista per il suo atteggiamento di indifferenza rispetto al sostrato biologico e materiale per l'identificazione e l'emulazione dei processi cognitivi<sup>109</sup>. Secondo i connessionisti, infatti, la costruzione di sistemi intelligenti non può prescindere dall'indagine sull'anatomia del cervello e sulle sue componenti: i neuroni<sup>110</sup>.

L'applicazione del pensiero connessionista ai sistemi di IA ha portato alla creazione di sistemi connessionistici che prendono il nome di *reti neurali*. Queste ultime riproducono il funzionamento dei neuroni, che sono alla base dei processi cognitivi animali<sup>111</sup>. Si sposta il fuoco della ricerca: dalla conoscenza intesa in senso logico e astratto, alle caratteristiche strutturali del cervello riprodotte in modo artificiale.

L'attenzione rivolta alle basi biologiche dell'intelligenza, ritenute essenziali per capirne la natura e per simularla su un calcolatore, spiega perché si suole riferirsi a questo filone della IA come IA di tipo biologico, in contrapposizione alla IA di tipo logico<sup>112</sup>.

In un agente che funziona attraverso reti neurali l'acquisizione della conoscenza avviene secondo il metodo *bottom-up*, cioè dal basso verso l'alto: partendo dall'osservazione di singoli dati, si giunge alla regola generale<sup>113</sup>. Ad esempio, se è affidato all'agente intelligente il compito di riconoscere le immagini che ritraggono il celebre quadro di Johannes Vermeer, *La ragazza col turbante*, esso analizza tutti i pixel delle immagini che costituiscono il *data set* e capirà da solo che nel quadro è rappresentata una figura con un turbante e un orecchino, ricavando la regola generale partendo dal particolare.

Esistono profonde differenze tra un sistema simbolista e connessionista, tanto in punto di rappresentazione, quanto in punto di acquisizione della conoscenza.

---

<sup>108</sup> voce *Regola di Hebb*, *Enc. della Scienza e della Tecnica*, Treccani, 2008, [www.treccani.it/enciclopedia/D\\_Hebb\\_L'organizzazione\\_del\\_comportamento](http://www.treccani.it/enciclopedia/D_Hebb_L'organizzazione_del_comportamento), Franco Angeli, 1975. La prima edizione del volume è del 1949. La regola di Hebb è sintetizzabile dall'espressione "*Fire together, wire together*", M. A. Boden, *L'intelligenza artificiale*, cit., p. 85.

<sup>109</sup> D. Marconi, *Linguaggio e scienza cognitiva*, cit., p. 201.

<sup>110</sup> A. Terrett, *Neural Networks*, cit., p. 95.

<sup>111</sup> D. Marconi, *Linguaggio e scienza cognitiva*, cit., p. 201.

<sup>112</sup> D. Parisi, voce *Intelligenza artificiale*, cit., [www.treccani.it](http://www.treccani.it).

<sup>113</sup> Voce *Bottom-up* in *Dizionario di economia e finanza*, Treccani, 2012, in [www.treccani.it](http://www.treccani.it)

Come già osservato, la rappresentazione della conoscenza, cioè il modo in cui la conoscenza è codificata per essere comprensibile al sistema, segue, nell'un caso, regole logiche, nell'altro, è il risultato delle connessioni che si instaurano tra i neuroni artificiali. Nella visione simbolista esiste una corrispondenza biunivoca tra i simboli e i concetti che rappresentano. Per converso, le rappresentazioni connessioniste sono distribuite, atteso che ogni dato della realtà è rappresentato attraverso le connessioni tra i neuroni, cosicché ogni neurone partecipa alla rappresentazione di concetti diversi.

Quanto al momento dell'acquisizione della conoscenza, cioè al modo in cui vengono selezionati i fatti rilevanti di un dato dominio, accade che l'apprendimento simbolista è consequenziale, cioè del tipo "se A, allora B". Nei modelli connessionisti l'apprendimento è parallelo, poiché tutti i neuroni imparano simultaneamente, secondo la regola di Hebb<sup>114</sup>.

Un'altra differenza risiede nel ruolo dell'uomo nella fase di elaborazione dei dati che costituiranno l'*input* per il modello di apprendimento automatico.

In base a questo rapporto con i dati, si è soliti delineare un'ulteriore distinzione, cioè quella tra modelli di apprendimento tradizionali e il *representation learning*<sup>115</sup>.

I modelli tradizionali richiedono l'intervento dell'uomo nella pre-elaborazione dei dati, poiché il dato grezzo è "incomprensibile" al sistema e deve essere tradotto in un linguaggio che sia a esso comprensibile. Il dato grezzo può essere rappresentato da un'immagine, dal testo di un documento, da un file audio e per essere compreso dal sistema di IA occorre che le sue caratteristiche siano trasposte in un codice. In questo modo dai dati sono estratte le caratteristiche o *features* di cui il sistema terrà conto. Si tratta di un compito altamente tecnico che richiede un lasso temporale non indifferente<sup>116</sup>.

Per contro, le tecniche che rientrano nel *representation learning* si caratterizzano per la capacità del sistema di elaborare dati grezzi. Ciò comporta che oggetto dell'analisi sono direttamente i dati e non le loro caratteristiche o *features* estrapolate da un esperto informatico. Per questo motivo i termini *feature* e *representation* sono intercambiabili<sup>117</sup>. Dunque, risulta ridotto il contributo dell'uomo nella pre-elaborazione dei dati.

Rientrano tra le tecniche di apprendimento tradizionali la regressione lineare, il *support vector machine* (SVM), gli alberi di decisione. Tra i modelli di *representation*

---

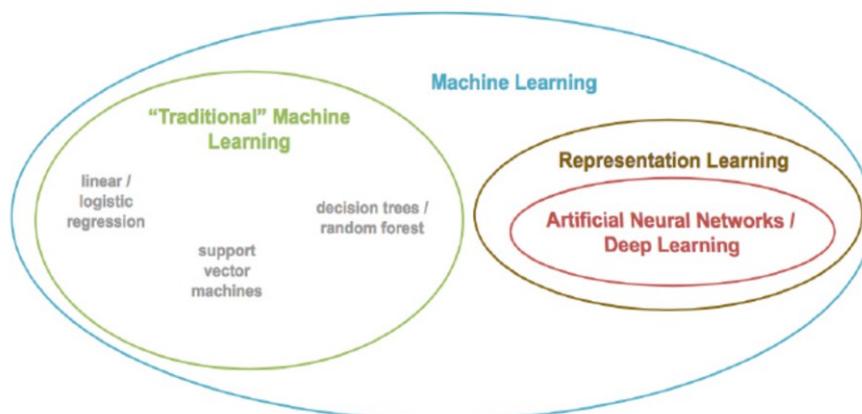
<sup>114</sup> P Domingos, *L'algoritmo definitivo*, cit., p. 122.

<sup>115</sup> J. Krohn, B. Grant, B. Aglaé, *Deep Learning Illustrated*, Addison-Wesley Professional, 2019, p. 27

<sup>116</sup> J. Krohn, B. Grant, B. Aglaé, *Deep Learning Illustrated*, Addison-Wesley Professional, 2019, p. 27

<sup>117</sup> *Id.* P. 27.

*learning*, i più diffusi sono le reti neurali, di cui un tipo particolare è rappresentato dall'apprendimento profondo o *deep learning*<sup>118</sup>.



**Immagine 3.1** – I diagrammi di Venn sintetizzano i rapporti tra i vari modelli di *machine learning*. L'immagine è tratta dal libro di J. Krohn, B. Grant, B. Aglaé, *Deep Learning Illustrated*, 2019.

La scelta del programmatore tra i vari modelli dipende dalla capacità dello stesso di adempiere al compito assegnato partendo dai dati che ha a disposizione.

Per agevolare il compito del programmatore sovviene la tecnica della *cross-validation*, con la quale si procede a un confronto tra vari modelli di *machine learning* al fine di scegliere il migliore in base al *task* affidato.

Di seguito analizzeremo, dapprima, un esempio di modello rientrante nella categoria del *traditional machine learning*, per poi soffermarci sulle reti neurali e sull'apprendimento profondo.

### 3.1 Alberi di decisione

L'apprendimento mediante alberi di decisione (*decision trees*)<sup>119</sup> applica il ragionamento logico mediante processi di inferenza induttiva. Esso è stato impiegato, tra l'altro, nell'ambito della diagnosi medica e nel settore del credito per calcolare il rischio di insolvenza<sup>120</sup>.

<sup>118</sup> J. Krohn, B. Grant, B. Aglaé, *Deep Learning Illustrated*, cit., p. 15.

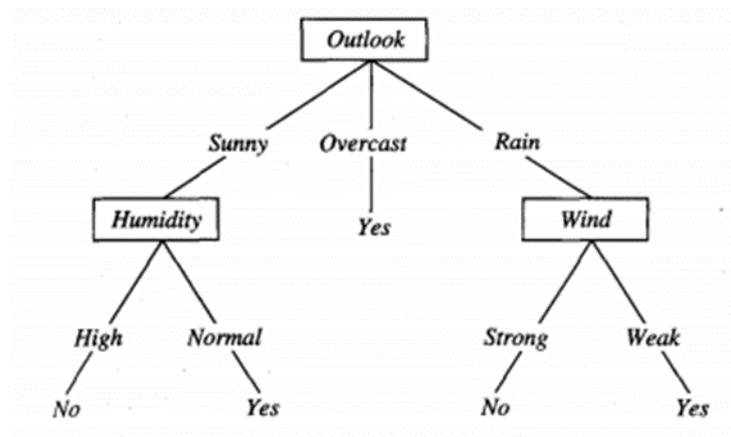
<sup>119</sup> N. J. Nilsson, *The Quest for Artificial Intelligence. A History of Ideas and Achievements*, Cambridge University Press, 2010, p. 500.

<sup>120</sup> T. Mitchell, *Machine Learning*, cit., p. 52.

In generale, un *albero* è formato da una serie di nodi collegati tra loro attraverso dei *link* o archi. I nodi contengono l'informazione e possono avere un solo arco di ingresso e più archi di uscita. Il primo nodo è detto *radice*. Da esso si dipana l'albero, le cui *foglie* saranno i nodi finali dai quali non fuoriescono altri archi. Il percorso che va dalla *radice* alle *foglie* è indicato come *path*.

L'albero è detto *di decisione* quando gli viene affidato il compito di estrarre un *output* partendo da un dato *input*. In tal caso, nel nodo sarà contenuta una domanda e da esso si dipaneranno due archi, uno con una risposta positiva e una negativa, contenute rispettivamente in due ulteriori nodi distinti. Le foglie conterranno la risposta finale.

Tale modello viene impiegato per la classificazione, di talché ogni foglia (cioè i nodi finali) rappresenta le classi predette. È altresì utilizzato per la regressione.



**Immagine 3.2.** Esempio di albero di decisione tratto dalla monografia di Tom Mitchell, riguardante l'individuazione della giornata ideale per giocare a tennis.

Tenendo in mente un caso di classificazione, la domanda posta all'interno di ciascun nodo concernerà una caratteristica dell'oggetto da classificare. Pertanto, nell'albero di decisione sono contenute istanze rappresentate da coppie 'attributo – valore'. Ad esempio, se volessimo chiederci quali sono le condizioni atmosferiche migliori per giocare a tennis, l'attributo sarà *outlook* (previsione) e i valori associati saranno *sunny*, *overcast*, *rain*. Come è dato vedere nell'immagine 3.2, ogni ramo uscente da un nodo corrisponde a uno dei possibili valori che l'attributo può assumere.

Per scegliere quale caratteristica inserire nel nodo, è dirimente il calcolo dell'entropia.

Il metodo di apprendimento degli alberi di decisione è di tipo *top down*. Nella parte iniziale dell'albero è contenuta una regola molto generale, che viene poi specializzata facendo crescere gradualmente l'albero.

Il modello di alberi di decisione presenta dei vantaggi non indifferenti. Anzitutto, esso è semplice da creare, ma, ancor più importante, riesce a spiegare come ha preso una certa soluzione<sup>121</sup>, poiché è possibile ricostruirne il percorso logico, e richiede poca attività di adeguamento dei valori. Questa è una caratteristica comune a tutti i modelli che seguono regole logiche secondo l'approccio simbolista della scienza cognitiva.

Tuttavia, in quanto modello rientrante nel *machine learning* tradizionale, pone il problema della necessaria pre-elaborazione dei dati, al fine di ricavarne le *features*, che costituiranno la rappresentazione della conoscenza per l'agente intelligente. Per questo, il sistema non è sempre utilizzabile, poiché in presenza di dati con un elevato numero di caratteristiche esso risulta molto lento<sup>122</sup>. Inoltre, è un modello soggetto al fenomeno dell'*overfitting*, come altri modelli, del resto<sup>123</sup>.

Il fenomeno del sovra-adattamento può essere evitato in due modi<sup>124</sup>: non facendo crescere l'albero, oppure lavorando sui dati ottenuti a seguito dell'*overfitting*.

Nel primo caso, si procede allo 'sfrondamento' dell'albero (*pruning*), rimuovendo quei nodi la cui presenza riduce la *accuracy* del *validation set*. L'attività di *pruning* continuerà fintantoché non risulti "dannosa", ad esempio nel caso in cui porti a una diminuzione dell'accuratezza rapportata al *validation set*<sup>125</sup>.

Quando non si dispone di una quantità di dati sufficienti da poter essere suddivisi in due insiemi (quello del *training* e quello della *validation*), si può fare ricorso alla regola del *post-pruning*. In questo caso, si consente di raggiungere una situazione di *overfitting* con riferimento all'insieme di dati usati nella fase dell'addestramento. L'esatta conoscenza dei dati è utilizzata per ricavare delle regole, che costituiranno la base di nuova conoscenza, senza tener conto del modo in cui si è giunti a esse. In altri termini, l'albero viene sfrondata dopo aver raggiunto l'*overfitting* in sede di addestramento e non

---

<sup>121</sup> R. Marmo, *Algoritmi per l'intelligenza artificiale*, cit., p. 279.

<sup>122</sup> Idem., p. 279.

<sup>123</sup> T. Mitchell, *Machine Learning*, cit., p. 68.

<sup>124</sup> Idem, p. 69 ss.

<sup>125</sup> Idem, p. 70.

prima, come accade in caso di *pruning* mero. Anche in questo caso, il *post-pruning* è interrotto laddove influisca negativamente sulla accuratezza, riducendola.

### 3.2 Reti neurali

Le reti neurali rappresentano un modello di apprendimento del filone della IA di tipo biologico, il cui studio è stato approfondito soltanto negli anni ottanta<sup>126</sup>. Questo modello riesce a ottenere ottimi risultati nell'interpretazione di dati di input complessi, come il riconoscimento di immagini in un sistema di guida autonoma<sup>127</sup>.

I sistemi basati sulle reti neurali sono in grado di operare su dati grezzi o incompleti e questa è la principale ragione del loro vasto impiego<sup>128</sup>. Esse, tuttavia, richiedono una grande quantità di dati, per questo la loro diffusione è stata accompagnata dall'aumento dei dati raccolti attraverso gli utenti di Internet.

Il modello trae ispirazione dalla struttura del cervello umano e, in particolare, dal neurone, l'unità fondamentale di tutto il tessuto del sistema nervoso.

Così come il neurone consta di un corpo cellulare, detto soma, da cui si ramifica un elevato numero di fibre chiamate dendridi e una singola fibra lunga detta assone, allo stesso modo, una rete neurale è composta un numero di nodi, o unità, connesse tramite dei collegamenti<sup>129</sup>.

Caratteristica fondamentale nelle reti neurali sono i pesi, cioè dei valori numerici associati a ogni singolo collegamento: è per il loro tramite che avviene l'apprendimento.

Vi sono unità connesse con l'esterno, attraverso le quali il sistema riceve gli *input*. Tali unità trasmettono il segnale a un livello di nodi intermedio (*hidden layer*), collegato a uno strato finale, il quale produrrà un *output*. Le unità nascoste non hanno connessioni dirette con l'esterno e la loro attività non può essere analizzata direttamente osservando il comportamento di input e output della rete<sup>130</sup>.

---

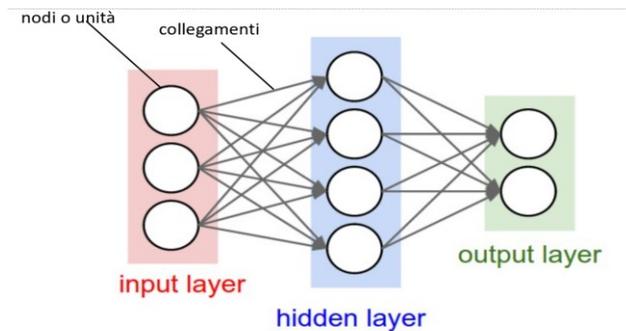
<sup>126</sup> D. Parisi, voce *Intelligenza artificiale*, cit., [www.treccani.it](http://www.treccani.it).

<sup>127</sup> T. Mitchell, *Machine learning*, cit., p. 81 e 83.

<sup>128</sup> A. Terrett, *Neural Networks - Towards Predictive Law Machines*, 3 *INT'L J.L. & INFO. TECH.* 94 (1995).

<sup>129</sup> Russel, Norvig, *Intelligenza artificiale*, cit., pp. 606 e 609.

<sup>130</sup> *Id.*, p. 614.



**Immagine 3.3.** Esempio grafico di una rete neurale con un solo strato nascosto.

L'idea è quella di una rete poiché ogni strato di nodi è connesso a quello precedente e ha un livello di attivazione. I vari nodi lavorano in parallelo e non in sequenza, come avviene nei sistemi di tipo logico. Inoltre, non vi è un controllo centralizzato<sup>131</sup>; ciò significa che la computazione è eseguita da più unità che lavorano senza un programma predefinito<sup>132</sup>.

Quanto alla fase di apprendimento, le prestazioni della rete non sono programmate dall'essere umano. Questi si limita a fissare le condizioni iniziali del processo di sviluppo e le regole da seguire durante tale processo. Spetterà al sistema elaborare da solo le sue prestazioni, attraverso un processo graduale di cambiamento, finché non è raggiunta la capacità desiderata<sup>133</sup>. Questo processo di cambiamento si attua attraverso i *pesi*, che, come vedremo, svolgono un ruolo fondamentale all'interno della rete.

### 3.2.1 *Perceptron*: come impara la rete

Le prime reti neurali erano reti semplici, chiamate *perceptrons* o perceptron, le quali non hanno unità nascoste<sup>134</sup>. A queste sono affidate compiti molto semplici, essendo molto limitate. Il perceptrone è il precursore delle reti neurali. Come già osservato, i modelli a reti neurali conobbero un declino negli anni sessanta e settanta a causa delle

<sup>131</sup> *Id.*, p. 638.

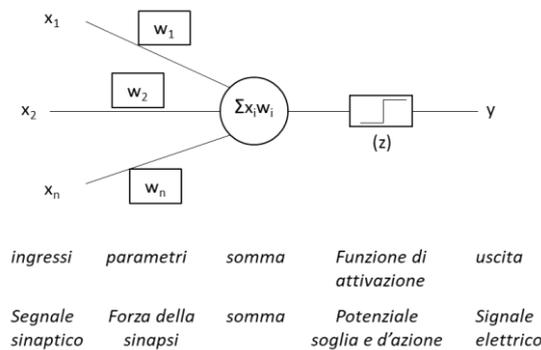
<sup>132</sup> A. Carobene, voce *Intelligenza artificiale – Approfondimento*, cit., [www.treccani.it](http://www.treccani.it).

<sup>133</sup> D. Parisi, voce *Intelligenza artificiale*, cit., [www.treccani.it](http://www.treccani.it).

<sup>134</sup> Le reti di perceptron furono create da Frank Rosenblatt, psicologo di Cornell, e il suo studio confluì nell'articolo *The Perceptron: a probabilistic model for information storage and organization in the brain*, pubblicato in *Psychological Review*, Vol. 65, n. 6, nel 1958. Il nome '*perceptron*' deriva dall'interesse di Rosenblatt verso le attività percettive del riconoscimento del linguaggio e della scrittura per le quali impiegò il suo modello di perceptrone.

critiche rivolte al perceptrone, in quanto incapace di riconoscere stimoli visivi anche molto semplici<sup>135</sup>.

Il neurone artificiale o *perceptron* ha un funzionamento simile al neurone biologico. I dati di *input* corrispondono al segnale sinaptico che dai dendriti è trasmesso al neurone. Vi sono poi dei parametri o pesi ( $w$ ), che indicano la “forza” del segnale di ingresso, cioè la sua importanza, attraverso un valore numerico. I pesi sono positivi o negativi e rappresentano, rispettivamente, connessioni eccitatorie e inibitorie<sup>136</sup>. I segnali di input, una volta raccolti, sono moltiplicati per il valore dei rispettivi pesi e, infine, ciascun risultato viene sommato. La somma dei prodotti dei pesi per i valori di ingresso costituisce la *funzione di attivazione*. Questa corrisponde al potenziale soglia o potenziale di azione del neurone biologico e svolge il compito di selezionare l’informazione ricevuta, determinando il valore del segnale in uscita  $Y$ . Tale informazione è trasmessa solo se è raggiunta una certa soglia prefissata. Se la soglia di attivazione è superata, l’informazione corrisponde al valore 1, mentre vale 0 nel caso contrario<sup>137</sup>.



**Immagine 3.4.** La funzione di attivazione  $f(z)$  determina il valore del segnale in uscita “Y” in base alla somma dei valori in ingresso ( $x_1, x_2, x_n$ ) moltiplicati per i relativi pesi ( $w_1, w_2, w_n$ ), ovvero “ $\sum x w$ ”. Vale pertanto la relazione  $Y = f(\sum x w)$ .

La caratteristica del perceptrone è che variando i pesi e le soglie è possibile modificare la funzione calcolata dal perceptrone: per decidere se il perceptrone si accenderà, basta

<sup>135</sup> È questa la critica di Marvin Minsky e Seymour Papert al perceptron di Rosenblatt, confluita nel volume *Perceptrons. An Introduction to Computational Geometry*, Mit Press, 1969.

<sup>136</sup> M.A. Boden, *L'intelligenza artificiale*, cit., p. 83.

<sup>137</sup> Il funzionamento di un perceptrone è descritto, tra gli altri, da P. Domingos, *L'algoritmo*, cit., p. 125, ma per la spiegazione matematica e tecnica del suo funzionamento si rinvia a P. Norvig, S. Russel, *Intelligenza artificiale*, cit., cap. 19, p. 605 ss., oltre che all'articolo dello stesso Rosenblatt.

aumentare o diminuire i pesi, a seconda della differenza che c'è tra il prodotto dei pesi per gli input e il valore di soglia.

Immaginiamo di voler domandare al sistema di imparare a distinguere le razze canine<sup>138</sup> e, tra queste, individuare quella del Golden retriever. Ogni figura di cane che sottoporremo al sistema costituisce un esempio, cui è attribuito un valore per ogni fattore. Se la somma dei pesi sarà pari o superiore al valore della soglia di attivazione, il sistema darà una risposta positiva, con uscita pari a 1; nell'esempio, pertanto, si tratterà di un cane Golden retriever e il sistema imparerà pian piano a riconoscerne le caratteristiche.

In questo caso, occorre fissare dei fattori che influenzino la decisione finale del sistema. Ad esempio, consideriamo la taglia dell'animale, il colore e lunghezza del manto. Ad ogni fattore corrisponderà un peso, il cui valore sarà tanto maggiore quanto più rilevante è il fattore. Nel caso del riconoscimento delle razze canine, un peso maggiore potrebbe essere dato alla taglia dell'animale (ad esempio, 0.7) e alla lunghezza del manto (ad esempio, 0.5), mentre un peso minore potrebbe essere quello corrispondente al colore del manto, non essendo questo un fattore sempre caratterizzante, posto che ogni razza ha di solito colori del manto diversi (ad esempio, 0.3).

**Pesi ( $w$ )**  
Taglia = 0.7  
Lunghezza manto = 0.5  
Colore manto = 0.3

Una volta determinati i pesi, vengono inseriti i segnali di ingresso o *input*. Questi saranno costituiti dai pixel di immagini di cani Golden retriever o dalla codifica di alcune caratteristiche. Ogni segnale di input ha un suo valore ( $x_1, x_2, \dots, x_n$ ). Per il Golden retriever, la caratteristica 'taglia grande' avrà il valore di 1, mentre 0 è il valore della caratteristica 'taglia piccola'. Per il colore del manto, se consideriamo il colore bianco, il valore sarà intermedio tra 0 e 1, essendo il manto solitamente mielato, da cui l'aggettivo 'golden' che dà il nome alla razza, ad esempio pari a 0.5. Consideriamo, infine, che un esempio di manto lungo abbia un valore di 0.8.

**Valori dati di input**  
Taglia grande = 1

---

<sup>138</sup> L'esempio è fatto, seppur con razze diverse da quella da noi scelta, da A. Terrett, *Neural Networks*, cit., p. 100.

Manto lungo = 0.8  
Colore bianco = 0.5

A questo punto si stabilisce il valore che deve essere raggiunto affinché il segnale sia trasmesso, cioè il valore minimo della soglia di attivazione, ad esempio 5.

Calcoliamo il valore della funzione di attivazione, moltiplicando ciascun input per il corrispondente peso che si è acceso e sommando questi valori:

$$\begin{aligned} f(z) &= (1*0,7) + (0,8*0,5) + (0,5*0,3) \\ &= 0,7+0,4+0,15 = 1,25 \end{aligned}$$

Si nota che il peso ‘taglia piccola’ non si sarà acceso, perché l’immagine ritrae un cane di taglia grande.

Nell’esempio, il valore della nostra funzione di attivazione è di 1,25. Essendo la soglia di attivazione pari a 5, il nostro perceptrone non sarà in grado di riconoscere la foto del Golden retriever e darà una risposta negativa. Ma il sistema sta errando, perché in foto è ritratto un cane Golden retriever. Ciò significa che la somma pesata dei suoi *input* dovrebbe essere più alta. Pertanto, bisogna aumentare i pesi degli input che si sono accesi.

L’esempio consente di capire perché l’apprendimento in una rete neurale avviene operando dei cambiamenti adattivi nei pesi<sup>139</sup>, cioè attraverso la progressiva modificazione del valore degli stessi. Più in generale, si osserva che la prestazione della rete dipende dai coefficienti quantitativi delle connessioni, che indicano la maggiore o minore capacità di trasmissione di ciascuna connessione<sup>140</sup>. Tali coefficienti sono inizialmente assegnati alla rete in modo casuale, poiché la rete non ha alcuna conoscenza iniziale<sup>141</sup> (nell’esempio, la rete non sa che la razza Golden retriever si caratterizza per avere una taglia grande e un manto lungo). I coefficienti vengono poi modificati dalla rete in base all’esperienza, cioè agli errori fatti.

Ancora una volta, saranno gli errori a guidare l’apprendimento: sbagliando, l’algoritmo impara.

---

<sup>139</sup> M. A. Boden, *L’intelligenza artificiale*, cit., p. 84.

<sup>140</sup> D. Parisi, voce *Intelligenza artificiale*, cit., [www.treccani.it](http://www.treccani.it).

<sup>141</sup> *Id.*, [www.treccani.it](http://www.treccani.it).

L'apprendimento potrà dirsi completato quando la rete trova un insieme di coefficienti che le consentono di rispondere correttamente ai dati di *input* segnalando l'*output* desiderato. Ciò dimostrerà che la rete ha acquisito una conoscenza che all'inizio non possedeva e che è rappresentata dall'insieme dei coefficienti trovati nel corso dell'apprendimento<sup>142</sup>.

Le reti con più unità nascoste sono dette reti multistrato e conoscono un largo impiego.

In base alle connessioni che si instaurano tra i diversi livelli, si distinguono reti con connessioni in avanti e con connessioni all'indietro.

Le prime sono le cosiddette reti *feedforward*, ove le connessioni tra i vari livelli seguono un andamento in avanti, appunto, cioè dal livello precedente a quello successivo, mentre non sono previste connessioni in senso opposto, cioè all'indietro o verso lo stesso livello<sup>143</sup>.

Le seconde sono le reti ricorrenti o *recurrent networks*, nelle quali le connessioni possono instaurarsi anche verso neuroni dello stesso livello o di livelli precedenti. Tali connessioni sono dette connessioni di feedback e presuppongono che l'informazione sia conservata al livello precedente per un certo arco di tempo. In questo modo, tali reti si caratterizzano per essere dotate di una memoria a breve termine<sup>144</sup>.

Le reti *feedforward* sono molto utilizzate. Ne costituiscono un esempio le reti neurali convoluzionali<sup>145</sup> impiegate nell'ambito del riconoscimento di immagini, di oggetti o di suoni sono utilizzate.

### **3.2.2 Deep learning: pregi e difetti**

Come già spiegato, le reti oggi utilizzate sono quelle multilivello. I molteplici livelli (*layers*) determinano la "profondità" o *depht* della rete. Per tale ragione, si è soliti definire

---

<sup>142</sup> *Id.*, [www.treccani.it](http://www.treccani.it).

<sup>143</sup> Per approfondimenti si rinvia a I. Goodfellow, Y. Bengio, A. Courville, *Deep learning* che dedicano alle deep feedforward networks l'intero cap. 6.

<sup>144</sup> *Id.*, cap. 10.

<sup>145</sup> R. Marmo, *Algoritmi per l'intelligenza artificiale*, cit., p. 349 ss. Per l'applicazione della tecnologia delle reti convoluzionali in ambito processuale, si rinvia a E. Athanasiadou, Z. Geradts & E. Van Eijk, *Camera Recognition with Deep Learning*, 3 *Forensic Sci. Res.* 210 (2018).

l'apprendimento attraverso le reti neurali con l'espressione *deep learning*<sup>146</sup>, mentre le reti neurali sono dette *reti neurali profonde* (*deep neural networks DNN*).

Il deep learning conosce una vasta applicazione in diversi settori, tra cui quello della visione artificiale. È stato notato che, negli anni, il livello di accuratezza raggiunto dai sistemi che adottano il deep learning è aumentato notevolmente, rispetto agli altri modelli. Questo dato è stato riscontrato nel corso della competizione americana ILSVRC (ImageNet Large Scale Visual Recognition Competition) che si svolge dal 2010 e in cui competono algoritmi di machine vision<sup>147</sup>. Nel 2012 fu presentato il primo algoritmo che impiegava la tecnologia del deep learning e si notò avere un livello di accuratezza nettamente maggiore. Dagli anni a seguire, la competizione è stata sempre vinta da sistemi di *deep learning* e nel 2017 un sistema di IA ha superato, quanto a accuratezza, l'uomo<sup>148</sup>.

Come già anticipato, una importante differenza tra le tecniche tradizionali di *machine learning* e il *deep learning* concerne la fase di pre-elaborazione dei dati.

Le tecniche tradizionali richiedono un lavoro di adattamento dei dati affinché possano costituire l'*input* del sistema di IA. Diversamente, il dato grezzo non sarebbe comprensibile dal sistema, il quale, pertanto, non riceverebbe alcun *input*. Questo processo è detto *feature engineering*<sup>149</sup>, è svolto dall'uomo e precede la fase di addestramento del modello.

Nel *deep learning* l'elaborazione del dato grezzo è affidata direttamente al sistema di IA e non richiede il contributo iniziale dell'uomo ovvero, se richiesto, questo è molto ridotto rispetto alle tecniche tradizionali di *machine learning*<sup>150</sup>.

Questo aspetto è fortemente caratterizzante il *deep learning* e le altre tecniche che rientrano nel *representation learning*. Tali modelli sono accomunati dalla capacità di apprendere le caratteristiche di input direttamente dai dati grezzi senza che sia necessario tradurli con un codice che sia compreso al sistema<sup>151</sup>.

Il dato grezzo, pertanto, costituisce l'*input* iniziale. A questo punto, il primo livello di rete neurale impara a rappresentare i dati con astrazioni semplici, mentre i livelli

---

<sup>146</sup> I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*, cit., p. 165.

<sup>147</sup> <http://image-net.org/challenges/LSVRC>

<sup>148</sup> *Machines 'beat humans' for a growing number of tasks*, in *Financial Times*, 30 novembre 2017

<sup>149</sup> J. Krohn, B. Grant, B. Aglaé, *Deep Learning Illustrated*, Addison-Wesley Professional, 2019, p. 14

<sup>150</sup> *Id.*, p. 15.

<sup>151</sup> *Id.*, p. 27. Nello stesso senso, Andrew Terrett, *Neural Networks - Towards Predictive Law Machines*, 3 INT'L J.L. & INFO. TECH. 94 (1995), p. 98.

successivi imparano a rappresentare astrazioni sempre più complesse rispetto a quelle del livello precedente.

In questo modo è drasticamente ridotta, se non del tutto eliminata, la fase di estrazione delle caratteristiche dei dati, che richiede anche un lavoro che dura anni se fatto manualmente da tecnici del settore.

Come nel singolo perceptrone, anche nell'ambito della rete un ruolo fondamentale è quello dei pesi. Il peso determina l'importanza del valore di input: se il dato in ingresso non è ritenuto importante, l'informazione non è trasmessa al livello successivo.

Le reti neurali sono addestrate usando un algoritmo di apprendimento induttivo<sup>152</sup>. Rientrando tra i modelli di *machine learning*, anche le reti neurali conoscono un apprendimento supervisionato o non supervisionato.

Il modello delle reti neurali presenta il vantaggio di riuscire a giungere a una decisione partendo da dati molto complessi e anche contenenti errori o rumori<sup>153</sup>.

Nel caso delle reti multistrato, tuttavia, la presenza di più livelli di nodi nascosti rende il modello più complesso rispetto al perceptrone e all'apprendimento che da esso discende. Mentre nel perceptrone si riesce a individuare l'errore fatto dai pesi, i quali possono essere modificati, tale operazione non può essere fatta nelle reti multilivello. Non esiste un metodo sicuro per modificare i pesi dei neuroni degli strati nascosti, così da ridurre gli errori commessi dallo strato di output<sup>154</sup>. Pertanto, mentre i modelli quali gli alberi di decisione riescono a spiegare il proprio ragionamento, le reti neurali “sono cumuli di numeri che nessuno può capire”<sup>155</sup>.

Il difetto del deep learning prende il nome di *black box*: gli strati profondi costituiscono una scatola nera cui non è possibile accedere per ricostruire il percorso che da un certo input ha condotto a un determinato output.

Il deep learning rappresenta un modello in cui forte è l'attrito tra capacità di predizione, da un lato, e spiegazione, dall'altro: è in grado di prevedere meglio il futuro rispetto a altri modelli, ma non si comprende la via seguita per realizzare la predizione. Si dispone solo di dati di input e output, e perciò la capacità esplicativa è ridotta al minimo<sup>156</sup>.

---

<sup>152</sup> Russel, Norvig, *Intelligenza artificiale*, cit., p. 609.

<sup>153</sup> T. Mitchell, *Machine learning*, cit., p. 85.

<sup>154</sup> P. Domingos, *L'algoritmo definitivo*, cit., p. 129.

<sup>155</sup> P. Domingos, *L'algoritmo definitivo*, cit., p. 149.

<sup>156</sup> A. Ruiz de Villa Robert e E. Puertas Prats, *I big data, il machine learning e la statistica*, volume II della collana *La matematica che trasforma il mondo*, RBA editore, 2020, p. 77.

Altro aspetto concerne i dati: le reti neurali richiedono una grande quantità di dati e una capacità computazionale elevata che possa processare una mole di dati notevole. Questi due fattori incidono sui soggetti che possono produrre sistemi di *deep learning*: l'elevata capacità computazionale e l'ingente volume di dati da gestire sono al di fuori della portata delle piccole aziende<sup>157</sup>. È un metodo che restringe la categoria dei produttori, creando grandi centri di potere. Sul piano della tutela dell'utente, ciò potrebbe scoraggiare le richieste risarcitorie per eventuali danni subiti dal sistema di IA, poiché il consumatore ben potrebbe rinunciare ad azionare la propria pretesa in giudizio qualora questa sia di scarso valore.

Una soluzione all'accentramento della capacità di produzione di sistemi di *deep learning* è rappresentata dall'apprendimento per trasferimento: i modelli già addestrati con un grande quantitativo di dati vengono riutilizzati su insiemi di dati più piccolo, per la soluzione di problemi più specialistici. Molte nuove aziende sfruttano grandi modelli già addestrati, specializzandoli in determinate aree<sup>158</sup>.

#### **4. Dati, Big Data, bias. L'algoritmo discriminatorio**

Nell'ambito del *machine learning* i dati assumono particolare rilevanza in quanto indispensabili per la creazione della base di conoscenza dell'algoritmo, il suo addestramento dell'algoritmo e per la valutazione del modello di *machine learning* prescelto. È quindi chiaro che l'algoritmo si 'nutre' di dati.

Il fatto che l'algoritmo abbia bisogno di dati impegna il giurista a occuparsi di questioni inerenti alle modalità con le quali i dati sono reperiti, ai criteri in base ai quali sono scelti dal programmatore o addestratore e, infine, al modo in cui vengono processati dall'algoritmo.

Il primo aspetto involge il tema della privacy e del trattamento dei dati personali, questione che non può essere approfondita in questa sede<sup>159</sup>.

Gli altri due punti, invece, conducono la riflessione al tema della discriminazione algoritmica, di cui di seguito ci occuperemo.

---

<sup>157</sup> A. Ruiz de Villa Robert e E. Puertas Prats, *I big data, il machine learning e la statistica*, volume II della collana *La matematica che trasforma il mondo*, RBA editore, 2020, p. 101.

<sup>158</sup> A. Ruiz de Villa Robert e E. Puertas Prats, cit., p. 101.

<sup>159</sup> Analizza il rapporto tra i dati personali e la rete, in generale, S. Rodotà, *Il modno nella rete. Quali i diritti, quali i vincoli*, Laterza, Roma-Bari, VI ed., 2019.

Il problema dell’algoritmo discriminatorio presuppone che a monte vi sia un algoritmo cui è demandato il compito di assumere delle decisioni che incidono su un soggetto o su una categoria di soggetti (*decisionmaking algorithm*) e che, nello svolgimento del suo compito decisionale, l’algoritmo possa assumere una decisione che sia discriminatoria.

Nel tentare di sistematizzare la problematica, le discriminazioni algoritmiche possono essere distinte in due categorie: da un lato, vi sono le discriminazioni digitali derivate, dall’altro le discriminazioni non intenzionali o discriminazioni digitali autonome<sup>160</sup>.

Le prime non sono proprie dell’algoritmo, ma vengono trasposte in esso dall’uomo. Questo atto del trasmettere una valutazione discriminatoria può essere intenzionale o involontario. La discriminazione potrebbe essere dovuta a una condotta tenuta dal programmatore nella fase della programmazione, ovvero in quella della configurazione del set di dati. Diverse sono le discriminazioni digitali autonome: esse derivano e dipendono dal funzionamento autonomo del sistema<sup>161</sup>.

Prendendo le mosse dalle discriminazioni digitali derivate, va notato che esse possono trovare origine nella fase di programmazione e di addestramento dell’algoritmo.

In sede di programmazione, è elaborato il codice che determinerà il funzionamento dell’algoritmo e che lo guiderà nella scelta. Tale codice potrebbe impiegare “categorie algoritmiche sospette”<sup>162</sup>, inserite dal programmatore volontariamente al fine di rifletterne i pregiudizi, ovvero immesse nel sistema a causa di un errore del programmatore medesimo<sup>163</sup>. Ciò può avvenire tanto al momento della definizione dell’obiettivo assegnato all’algoritmo, che nella scelta delle variabili o delle categorie in base alle quali è costruita la decisione algoritmica<sup>164</sup> e di cui l’algoritmo dovrà tenere conto al momento della scelta. Infatti, ai programmatori è demandato il delicato compito di tradurre in linguaggio informatico l’obiettivo assegnato all’algoritmo, il quale può

---

<sup>160</sup> La terminologia qui adoperata riprende, in parte, quella impiegata da A. Venanzoni, *La valle del perturbante: il costituzionalismo alla prova delle intelligenze artificiali e della robotica*, in *Politica del diritto*, 2/2019, p. 253, il quale distingue tra pregiudizi digitali derivati e pregiudizi digitali autonomi.

<sup>161</sup> Questa distinzione è tracciata in modo velato da S. Barocas, A. D. Selbst, *Big Data’s Disparate Impact*, in *California Law Review*, Vol. 104, pp. 671 ss., e ripresa anche da P. Zuddas, *Intelligenza artificiale e discriminazioni, Liber amicorum per Pasquale Costanzo*, in *Consulta Online*, 16 marzo 2020, p. 5.

<sup>162</sup> D.U. Galetta, J.G. Corvalán, *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *Federalismi.it*, n. 3/2019, 21-22.

<sup>163</sup> S. Barocas, A. D. Selbst, *Big Data’s Disparate Impact*, cit., p. 678.

<sup>164</sup> G. Resta, *Governare l’innovazione tecnologica*, cit., p. 217.

presentare delle caratteristiche oggettive chiare, ovvero può presupporre una valutazione. Ad esempio, mentre è maggiormente ispirata a criteri oggettivi la distinzione tra cosa è spam e cosa non lo è, non è tale la nozione di “buon dipendente”<sup>165</sup>. Quest’ultimo concetto può essere definito ricorrendo a caratteristiche diverse e, tra queste, potrebbe essere preso in considerazione un elemento discriminatorio<sup>166</sup>. Si pensi, ad esempio, ai concetti di razza, sesso, alla provenienza da una certa area geografica, alla condizione economica. In questo modo, trova spazio nel sistema una discriminazione derivata, cioè “ereditata” dal programmatore, anche se non intenzionalmente. Ne deriva che la discriminazione algoritmica si collocherà solo a valle di una discriminazione a monte, sicché l’elaborazione dei dati sarà condizionata da una condotta discriminatoria umana.

Di recente, proprio l’esempio in esame del “buon lavoratore” è stato oggetto di una ordinanza adottata dal Tribunale di Bologna nel dicembre 2020<sup>167</sup> con la quale è stata accertata e dichiarata la discriminatorietà della condotta della società Deliveroo, in relazione alle condizioni di accesso alla prenotazione delle sessioni di lavoro da parte dei lavoratori (i *riders*) che si occupano della consegna di cibo a domicilio. Costoro ottengono un punteggio dalla piattaforma, calibrato alla luce di due parametri, l’indice di affidabilità e l’indice di partecipazione nei picchi. Più alto è il punteggio del *rider*, maggiore sarà la possibilità di procacciarsi occasioni di lavoro, poiché chi ha il punteggio più alto può segnalare alla piattaforma la propria disponibilità scegliendo con priorità rispetto agli altri la fascia oraria in cui effettuare la consegna, privando di occasioni di lavoro i *riders* con il punteggio più basso. Quanto ai due parametri, l’indice di affidabilità tiene conto del “numero delle occasioni in cui il rider, pur avendo prenotato una sessione, non ha partecipato, dove “partecipare” significa loggarsi entro i primi 15 minuti dall’inizio della sessione”; mentre, l’indice di partecipazione nei picchi considera “il numero di volte in cui ci si rende disponibili per gli orari (dalle ore 20 alle ore 22 dal venerdì alla domenica) più rilevanti per il consumo di cibo a domicilio”. Il Tribunale ha ritenuto che i due parametri realizzassero un effetto discriminatorio poiché sarebbe indirettamente ostacolato l’esercizio del diritto di sciopero da parte del *rider*. Questi, infatti, per aderire

---

<sup>165</sup> S. Barocas, A. D. Selbst, *Big Data’s Disparate Impact*, cit., p. 678 – 679.

<sup>166</sup> Sarebbe tale secondo G. Resta il criterio della puntualità nel recarsi sul luogo di lavoro, in quanto ciò può finire per penalizzare sistematicamente tutti coloro che vivono in periferia. “E poiché in determinati contesti la circostanza dell’abitare in periferia è una variabile strettamente correlata alle origini etniche e alle condizioni sociali più disagiate, un siffatto criterio formalmente «neutro» potrebbe finire per riflettersi negativamente a danno di categorie già svantaggiate”, cit. p. 217.

<sup>167</sup> Trib. Bologna, sezione Lavoro, causa r.g. 2949/2019, 31 dicembre 2020, *CGIL Bologna contro Deliveroo s.r.l.*

a uno sciopero indetto in un giorno o fascia oraria in cui abbia dato la sua disponibilità a lavorare, dovrebbe non partecipare alla sessione prenotata -e non preventivamente cancellata- e ciò lo esporrebbe a una valutazione negativa da parte dell'algoritmo, poiché calerebbe il parametro della affidabilità. Il Tribunale ha dipoi ritenuto che anche una cancellazione in ritardo della prenotazione da parte del *rider* (cioè una cancellazione che non avvenga entro le 24 ore precedenti l'inizio della sessione prenotata) lo penalizzerebbe, esponendolo a un punteggio negativo.

Ancora, l'algoritmo non tiene conto delle ragioni per cui la cancellazione è effettuata dal lavoratore (ad esempio, motivi di salute) e, secondo il Tribunale, considerare irrilevanti i motivi della mancata partecipazione alla sessione prenotata o della cancellazione tardiva implicherebbe riservare lo stesso trattamento a situazioni diverse, così realizzandosi una discriminazione indiretta.

Quello sottoposto all'attenzione del Tribunale bolognese è un caso di algoritmo considerato discriminatorio per via delle variabili che orientano la decisione del sistema e che ne condizionano la operatività. Ma vi sono altri modi in cui può realizzarsi una discriminazione digitale derivata.

In particolare, l'effetto discriminatorio può verificarsi anche allorché il codice di decisione sia orientato sulla base di categorie solo apparentemente neutre, ma che di norma ricorrono con frequenza nelle categorie protette<sup>168</sup>. In questo caso, la discriminazione algoritmica è frutto di *proxies*. La proxy discrimination è più difficile da sventare poiché il dato utilizzato è formalmente neutro, ma il suo impiego in un certo contesto produce un effetto discriminatorio. È questo il caso in cui un istituto di credito utilizzi ai fini della profilazione dei propri clienti il codice postale dell'indirizzo di residenza. Sebbene si tratti di un dato formalmente non discriminatorio, discriminante sarà il suo impiego, poiché potrebbero essere fatte delle distinzioni tra i soggetti che abitano nel centro della città e chi abita in periferia, se il sistema incrocerà il dato del luogo di residenza con il livello di potenziale insolubilità del cliente, così creando dei *clusters* discriminanti.

Stiamo ora analizzando le proxy discrimination nel contesto delle discriminazioni digitali derivate poiché le proxies possono essere di origine umana: in altri termini, è possibile che chi programma l'algoritmo ordini al sistema di selezionare o escludere

---

<sup>168</sup> P. Zuddas, *cit.*, p. 6.

soggetti appartenenti a categorie protette individuate attraverso un carattere proxy. Questi sono casi facilmente individuabili e assoggettabili a un controllo preventivo. Maggiori problemi pongono le *proxy discriminations* di origine algoritmica<sup>169</sup>. Si tratta dei casi in cui le correlazioni attraverso caratteri proxy sono effettuate autonomamente dall'algoritmo, senza che ciò sia stato previsto in sede di programmazione<sup>170</sup>. Questo tipo di discriminazione è più difficile da sventare: bisognerebbe provare che il sistema ha ricavato dai dati di input dei *pattern* che penalizzano una certa categoria a discapito di un'altra. In ogni caso, va detto che si tratta di una discriminazione appartenente alla seconda categoria, cioè alle discriminazioni non intenzionali o digitali autonome.

Volendo però ultimare l'esame delle discriminazioni digitali derivate, prima di passare all'esame di quelle autonome, bisogna sottolineare che l'intento discriminatorio può realizzarsi non solo nella fase di programmazione dell'algoritmo, ma anche nella successiva fase di addestramento dello stesso. In questo caso, è fondamentale la scelta dell'insieme di dati da sottoporre all'algoritmo. Se i dati di input forniscono una conoscenza parziale o orientata al sistema, si verificherà un problema di *bias*, cioè di errore sistematico.

In tale evenienza, l'agente umano responsabile va individuato in colui il quale è tenuto a scegliere e costituire l'insieme di dati di input. Se tale soggetto è diverso dall'addestratore, è ragionevole ritenere che anche questi, cioè l'addestratore, potrà essere considerato responsabile della discriminazione, quantomeno in via colposa. Difatti, in sede di addestramento dell'algoritmo, egli ben avrebbe potuto notare l'errore sistematico o *bias* e perciò potrebbe essere chiamato a rispondere per la scarsa perizia con la quale ha svolto la mansione a lui affidata.

Potrebbe sostenersi la natura autonoma della discriminazione che trova origine nel set di dati, argomentando che la discriminazione discende dall'interpretazione dei dati da parte del sistema. L'obiezione non sarebbe però accoglibile. Infatti, l'interpretazione ha ad oggetto il contenuto dei dati scelti in fase di addestramento. E allora, nel caso dell'interpretazione artificiale, il problema risiede non tanto nell'atto dell'interpretare in sé considerato, quanto nella selezione dei dati inseriti nel sistema. Quest'ultima operazione è fatta dall'uomo e consiste, in fondo, nell'interpretazione di una data realtà,

---

<sup>169</sup> P. Zuddas, *cit.*, p. 7.

<sup>170</sup> L'argomento è ampiamente approfondito da A. Prince, D.B. Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, in *Iowa Law Review*, 5 agosto 2019.

circoscritta in un insieme (di dati): ad esempio, il concetto di “cane” è dato dall’insieme delle immagini di cani selezionate dall’uomo, così come il concetto di pericolosità sociale è desunto dall’insieme di informazioni prescelte dall’uomo. Una volta formato il set di dati, esso costituirà poi il “testo” sottoposto all’interpretazione artificiale. Questo ci aiuta a collocare l’errore che conduce alla discriminazione in un momento ben preciso, che è quello della scelta dei dati.

La scelta dei dati consiste nel primo momento interpretativo e tale prima interpretazione è demandata all’uomo. A essa segue la seconda interpretazione, effettuata dalla macchina o sistema intelligente, il quale giungerà a una decisione interpretando l’interpretazione umana.

Dove risiede allora il problema? Il problema si colloca al momento della scelta dei dati. Secondo Gadamer, l’interpretazione muove sempre da una pre-comprensione, da un pre-giudizio sul senso complessivo del testo da interpretare<sup>171</sup> e se, nella prospettiva qui descritta, il “testo da interpretare” consiste anzitutto nella realtà interpretata dall’uomo attraverso i dati, allora il pre-giudizio cui si riferisce Gadamer è essenzialmente un pregiudizio umano.

L’interpretazione della macchina, partendo dal “testo” scritto dall’uomo – “testo” che consiste nei dati -, è un’interpretazione già orientata dalla precedente interpretazione umana.

Questo ci consente di precisare alcuni aspetti del nostro discorso. Se “interpretare” vuol dire *“intendere e spiegare nel suo vero significato (o in quello che si ritiene sia il significato giusto o più probabile) il pensiero d’uno scritto o d’un discorso il cui senso sia oscuro o dia luogo a dubbi”*<sup>172</sup>, allora questa attività deve considerarsi attività propria dell’uomo. La macchina non è in grado di capire il senso della domanda che le si pone e, di conseguenza, del “testo” (cioè dell’insieme dei dati) che è chiamata a interpretare. Piuttosto essa si limita a tradurre i dati, cioè a trasporli in un altro codice linguistico. L’uomo interpreta, la macchina traduce. La devianza dalla neutralità starà nel momento interpretativo, in quanto lì risiede una valutazione, e non nel momento della traduzione, in quanto atto “meccanico”.

Se ci si interroga sulla neutralità dei sistemi di IA, potremmo dire che la neutralità è di fatto irraggiungibile perché la neutralità non appartiene all’uomo. Ciò è in linea con

---

<sup>171</sup> H. D. Gadamer, *Verità e metodo*, Bompiani, Milano, 1983, pp. 312 ss.

<sup>172</sup> <https://www.treccani.it/vocabolario/interpretare>

l'idea per cui il computer (e come esso qualsiasi sistema intelligente) non ha una logica propria, ma è dotato di strutture che applicano qualsiasi logica, in particolare, la logica impostagli dal programma, cioè dall'uomo che lo ha formulato<sup>173</sup>.

L'insieme di dati da cui prende le mosse il ragionamento algoritmo può essere di due tipi: "chiuso", cioè definito in modo stabile, ovvero "aperto", cioè non determinabile aprioristicamente. Nei sistemi aperti l'insieme dei dati di input è difficile da controllare, per questo la responsabilità del programmatore o dell'addestratore non sarà semplice da dimostrare. In questo caso, infatti, l'esito discriminatorio può non discendere dall'opera dell'uomo, bensì essere autonomo.

Ciò ci consente di introdurre la seconda tipologia di discriminazioni, cioè le discriminazioni non intenzionali o discriminazioni digitali autonome, derivanti dal modo in cui l'algoritmo opera nel mondo reale.

Il problema si pone soprattutto per i sistemi di *machine learning* collegati a una fonte open source di dati. Ci riferiamo ai casi in cui l'algoritmo non presenta problemi nel set originario di dati, e quindi la discriminazione non trova origine al momento dell'addestramento, ma continua la fase di apprendimento in modo autonomo prendendo i dati da Internet, i quali possono essere altamente discriminatori, poiché spesso riflettono i pregiudizi della società<sup>174</sup>. Un esempio, già analizzato, è quello delle *proxies discriminations* di origine algoritmica.

Il sistema di *machine learning* può anche modificare il proprio funzionamento alla luce delle indicazioni ricevute dagli utilizzatori<sup>175</sup>, riproducendone le tendenze discriminatorie.

---

<sup>173</sup> R. Borruso, *Analisi giuridica del computer*, in *Computer e diritto*, tomo I, Giuffrè, Milano, 1988, p. 101. Sul problema, specifico, della neutralità delle scelte e delle decisioni del motore di ricerca, si veda il lavoro di G. Fioriglio, *La "dittatura" dell'algoritmo: i motori di ricerca web e neutralità della indicizzazione. Profili informatico-giuridici*, in *Bocconi Legal Papers*, 5, 2015, p. 113-140. *HeinOnline*.

<sup>174</sup> È noto alle cronache il caso di Tay, chatbot creato da Microsoft perché operasse su Twitter e su altre piattaforme social con l'obiettivo di avviare e sostenere conversazioni con giovani americani tra i 18 e i 24 anni. Il chatbot ha in breve tempo mutuato dalla rete commetti razzisti, sessisti e xenofobi e per questo è stato censurato dalla stessa Microsoft e non più utilizzato. <https://www.lastampa.it/tecnologia/news/2016/03/25/news/tweet-razzisti-microsoft-chiude-il-chatbot-tay-1.36583608>.

<sup>175</sup> A. Chander, *The Racist Algorithm?*, in *Michigan Law Review*, April 2017, p. 1034 ss., che cita il caso del sessismo riscontrato nell'operazione di completamento automatico delle parole inserite nel motore di ricerca Google. Una pubblicità promossa dell'ONU nel 2013 mostrava che la ricerca effettuata su Google che iniziasse con le parole "*women should*" era seguita da frasi sessiste per effetto dell'operazione di completamento automatico, che prendeva in considerazione le ricerche precedenti effettuate con maggiore

In questi casi, pertanto, la discriminazione è autonoma, frutto del momento di apprendimento automatico da parte del sistema e può non essere prevedibile dai programmatori e dagli addestratori nel momento in cui il software viene immesso nel mercato. Il sistema non presenta errori o manomissioni né nella fase della programmazione, né in quella dell'addestramento e, tuttavia, perviene nel corso del suo utilizzo a decisioni discriminatorie. Se la causa della discriminazione non è individuabile, si pone il problema del *black box causale*, che rende difficoltosa l'individuazione di un soggetto responsabile per via dell'opacità del sistema. Ciò può condurre a risultati arbitrari, qualificabili come tali solo a seguito di una valutazione cui lo stesso algoritmo deve essere sottoposto, ad esempio, comparando il risultato cui è pervenuto con i risultati di sistemi simili chiamati a analizzare contesti analoghi<sup>176</sup>.

#### 4.1 Dalla neutralità alla esplicabilità della decisione algoritmica

Oltre agli scenari sin qui analizzati, va rilevato che gli algoritmi decisionali basati sui dati presentano dei rischi "intrinseci", in quanto connessi al funzionamento stesso di questo tipo di algoritmi.

Anzitutto, bisogna considerare che un problema endemico dei *big data* è che vi sono categorie di soggetti la cui profilazione non è possibile poiché costoro non hanno accesso alla rete. Si tratta di *people who live on big data's margins*<sup>177</sup>, cioè ai margini di quella società artificiale creata dai dati estratti dalla rete. Ciò comporta che anche un insieme di dati correttamente formato può essere *statistical biased*<sup>178</sup> se fallisce nel rappresentare un certo gruppo in modo proporzionato rispetto a altro gruppo.

Altro rischio intrinseco è costituito dai limiti imposti dal modello statistico. La complessità della società, con le peculiarità di ciascun individuo o classe di individui non può in alcun modo essere colta da un modello statistico. Per definizione, il modello statistico deve selezionare solo alcune caratteristiche e tralasciarne delle altre. In questo

---

frequenza dagli utenti. *UN Women Ad Series Reveals Widespread Sexism*, UN WOMEN (Oct. 21, 2013), <http://www.unwomen.org/en/news/stories/2013/10/women-should-ads> [<https://perma.cc/L9ZU-RX2G>].

<sup>176</sup> L'idea di confrontare i risultati forniti dagli algoritmi impiegati in contesti simili o identici è stata messa in atto nello studio dei sistemi di *credit scoring*, impiegati dagli istituti di credito al momento della concessione di mutui o finanziamenti a nuovi clienti. Attraverso una comparazione dei sistemi di scoring utilizzati da tre diversi istituti di credito, è stato dimostrato che al 29% dei clienti era stato attribuito un punteggio diverso, con un margine di almeno 50 punti di differenza. C. Carter et al., *The Credit Card Market and Regulation: In Need of Repair*, 10 *N.C. Banking inst.* 23, 2006, p. 41.

<sup>177</sup> J. Lerman, *Big Data and its Exclusions*, 66, *Stanford Law Review online*, 2013, pp. 55, 57.

<sup>178</sup> S. Barocas, A. D. Selbst, *Big Data's Disparate Impact*, cit., p. 684.

caso la discriminazione è definita da taluni come “razionale”, in quanto è il risultato obbligato del ragionamento statistico<sup>179</sup>.

Nell’ambito degli algoritmi predittivi, cioè quelli deputati a fare delle previsioni future, tali rischi intrinseci sono accentuati dalla propensione del sistema a “codificare il passato”, cristallizzando un determinato “stato del mondo” nel processo prognostico<sup>180</sup>. Questa tecnica è vantaggiosa quando si prendono in esame accadimenti che non involgono operazioni valutative, bensì esclusivamente scientifiche, come, ad esempio, l’andamento delle perturbazioni per prevedere le future condizioni meteorologiche<sup>181</sup>. Lo stesso processo, tuttavia, assume connotati di problematicità quando tali valutazioni sono riferite a stati o condizioni dell’uomo, ovvero a processi sociali e, in particolare, quando l’algoritmo predittivo è chiamato a esprimere un giudizio che ricada su un individuo o una categoria di individui<sup>182</sup>.

Sono algoritmi predittivi, ad esempio, quelli impiegati nell’ambito della giustizia penale per valutare il rischio di reiterazione del reato, tra questi il sistema COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*)<sup>183</sup> o, nell’ambito della giustizia minorile, il sistema SAVRY (*Structured Assessment of Violence Risk in Youth*)<sup>184</sup>.

---

<sup>179</sup> “By definition, data mining is always a form of statistical (and therefore seemingly rational) discrimination”. In questo senso si esprimono S. Barocas, A. D. Selbst, *Big Data’s Disparate Impact*, cit., p. 677 e p. 688.

<sup>180</sup> G. Resta, *Governare l’innovazione tecnologica*, cit., p. 214.

<sup>181</sup> G. Resta, *Governare l’innovazione tecnologica*, cit., p. 214.

<sup>182</sup> D. Keats Citron e F. Pasquale, *The Scored Society: Due Process for Automated Predictions*, in *University of Maryland Francis King Carey School of Law, Legal Studies Research Paper*, cit., descrivono lo scenario creato dai sistemi di *scoring*, in cui le decisioni vengono assunte alla luce di un sistema di punteggio attribuito da giudizi espressi in passato talvolta anche dall’utente umano attraverso il noto sistema delle recensioni.

<sup>183</sup> Algoritmo che prende in considerazione le risposte date a un questionario di 137 domande, divise in cinque macro gruppi: «criminal involvement, relationships/lifestyles, personality/attitudes, family, and social exclusion». Il sistema COMPAS è stato e continua a essere oggetto di molte riflessioni da parte dei giuristi, stante la sua denunciata tendenza discriminatoria a danno dei soggetti di colore, notata per prima dall’associazione no profit PRO PUBBLICA. J. Angwin, J. Larson, S. Mattu, L. Kirchner, *Machine Bias*, in [www.propublica.org](http://www.propublica.org), 23 maggio 2016. Per approfondimenti, si rinvia a M. Gialuz, *Quando la giustizia penale incontra l’intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, *Dir. Pen. Cont.*, in [www.penalecontemporaneo.it](http://www.penalecontemporaneo.it); G. Ubertis, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Sistema penale 2020*. Sull’impiego di questi strumenti nel nostro ordinamento A.M. Maugeri, *L’uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra evidence based practices e tutela dei diritti fondamentali*, in *Arch. Pen.*, 1, 2021, pp. 1-37.

<sup>184</sup> Il sistema SAVRY rientra tra gli *Structured Professional Judgment* (SPJ), a supporto della valutazione condotta da esperti e lascia loro un margine decisionale, quindi è un sistema aperto e interpretabile, a differenza di COMPAS cfr. S. Tola, M. Miron, E. Gómez, C. Castillo, *Why machine learning may lead to unfairness: evidence form risk assessment for Juvenile Justice in Catalonia*, Best Paper Award, International Conference on AI and Law, 2019. Gli strumenti impiegati per la prevenzione del crimine minorile si distinguono in strumenti per individuare il rischio di commettere reati da parte di soggetti che

Proprio quello che qui definiamo *rischio intrinseco* della decisione algoritmica ha portato alcuni studiosi a rinunciare al concetto di neutralità dell'algoritmo<sup>185</sup>. La necessità di servirsi di dati che non possono che essere sempre parziali conduce a una situazione di *bias inevitabile*, poiché talvolta è impossibile fornire una rappresentazione neutra dell'informazione iniziale.

Si nota, inoltre, che è vano esigere la neutralità da parte di un sistema progettato per scegliere, filtrare o ordinare le informazioni secondo certi principi e concepito proprio al fine di ridurre l'incertezza in un universo in cui l'abbondanza di dati non consente di scegliere<sup>186</sup>.

Per questo, è stato sostenuto che, a fronte di una *vana neutralità*, è più corretto esigere la *lealtà dell'algoritmo*<sup>187</sup>. Si tratta di richiedere al programmatore e a coloro che mettono l'algoritmo a disposizione di terzi di chiarire i criteri che presidono la decisione algoritmica e garantirne la verificabilità, affinché possa essere chiaro il funzionamento del sistema, così da evitare che la parzialità della conoscenza su cui si basa la decisione possa celare una discriminazione non identificabile. Si tratta della caratteristica dell'*esplicabilità (explicability)*<sup>188</sup>, che deve connotare i sistemi di IA intesi come sistemi che rispondono al requisito della trasparenza. La neutralità algoritmica si realizza con un impegno sul piano della *trasparenza* circa gli algoritmi utilizzati, il loro funzionamento e la loro implementazione, consentendo l'accesso alle fasi del ragionamento seguite dal sistema per assumere una determinata decisione<sup>189</sup>.

---

non abbiano ancora alcun precedente e strumenti per individuare il rischio di recidiva, cfr. J. Savignac, *Tools to Identify and Assess the Risk of Offending Among Youth*, Published by National Crime Prevention Centre (NCPC), 2010, che offre un quadro sugli strumenti adoperati nel sistema canadese.

<sup>185</sup> D. Cardon, *Le Pouvoir des algorithmes*, cit., p. 65.

<sup>186</sup> D. Cardon, *Le Pouvoir des algorithmes*, cit., p. 65.

<sup>187</sup> Idem, p. 66; l'opinione è anche quella del Conseil d'État, *Le Numérique et les droits fondamentaux*, Paris, *La Documentation française*, 2014. Secondo l'organo di giustizia, «*Les plateformes devraient être soumises à une obligation de loyauté envers leurs utilisateurs, tant les utilisateurs non professionnels dans le cadre du droit de la consommation que les utilisateurs professionnels dans le cadre du droit de la concurrence*», p. 21.

<sup>188</sup> «*Tale principio implica che i processi devono essere trasparenti, le capacità e lo scopo dei sistemi di IA devono essere comunicati apertamente e le decisioni, per quanto possibile, devono poter essere spiegate a coloro che ne sono direttamente o indirettamente interessati*», ed è di uno dei principi etici elaborati dal Gruppo di esperti ad alto livello sull'intelligenza artificiale nel documento *Orientamenti Etici per un'IA affidabile*, p. 14 I principi o imperativi etici sono quattro: i) il rispetto dell'autonomia umana, ii) la prevenzione dei danni, iii) l'equità, iv) l'esplicabilità.

<sup>189</sup> È la prospettiva sostenuta anche da G. Fioriglio, *La "dittatura" dell'algoritmo*, cit., p. 137, con riguardo ai software utilizzati dai motori di ricerca. Secondo l'A., sussisterebbe un generale obbligo di indicizzazione corretta in capo a colui che presta il servizio di motore di ricerca e dichiara che tale servizio sia reso in modo

Il sistema di IA risponderà al criterio della esplicabilità se il produttore è in grado di anticipare gli effetti cui può condurre l'utilizzo del sistema, in modo tale che l'utilizzatore possa per tempo identificarli e, se del caso, segnalare eventuali anomalie<sup>190</sup>. Naturalmente, il livello di trasparenza richiesto aumenta allorché l'algoritmo decisionale sia in grado di incidere valori o diritti fondamentali<sup>191</sup>.

Pertanto, il produttore è destinatario di un obbligo di lealtà nella produzione, la cui giuridicizzazione è possibile attraverso il richiamo all'obbligo contrattuale di correttezza e buona fede che dà corpo all'obbligo di informazione del professionista nei confronti del consumatore. In caso di inadempimento di tale obbligo, il rimedio previsto sul piano civilistico avrà natura contrattuale e risarcitoria.

Sul piano penale, l'inadempimento dell'obbligazione potrà integrare, anzitutto, gli estremi del reato di truffa contrattuale, oltre a potersi configurare la possibilità di un concorso di reati nel caso di lesione di beni giuridici ulteriori rispetto a quelli tutelati dal reato di truffa.

Rendendo trasparente il funzionamento dell'algoritmo è possibile esercitare un controllo sulla programmazione ed è possibile eliminare la discriminazione poiché è noto il modo in cui opera l'algoritmo.

Questo, ad esempio, è quanto si è verificato nel caso Deliveroo portato all'attenzione del Tribunale di Bologna, sezione Lavoro.

Quando, però, la discriminazione è riconducibile ai dati e l'algoritmo replica le discriminazioni in essi contenuti a causa del metodo statistico, la *transparency by design* o *explicability* potrebbe non essere sufficiente. Il sistema, infatti, non presenterebbe errori nel modo in cui funziona e la discriminazione sarebbe il riflesso delle disuguaglianze che

---

neutrale e non discriminatorio. Inoltre, sempre l'A. precisa che tale obbligo acquista una cogenza maggiore allorché il prestatore del servizio si trovi in una posizione dominante sul mercato. Cfr. pp. 138 e 139.

<sup>190</sup> In questo senso, D. Cardon, *Le Puovoir*, cit., p. 70.

<sup>191</sup> Come sottolineato dal Gruppo di esperti ad alto livello sull'intelligenza artificiale nel documento *Orientamenti Etici per un'IA affidabile*, ove si legge che "Il grado di esplicabilità necessario dipende in larga misura dal contesto e dalla gravità delle conseguenze nel caso in cui il risultato sia errato o comunque impreciso", p. 15, "Ad esempio, se le raccomandazioni di acquisto generate da un sistema di IA sono imprecise ciò non suscita grandi preoccupazioni etiche, mentre diversa è la situazione quando i sistemi di IA devono valutare l'opportunità o meno di accordare a una persona condannata a pena detentiva la libertà condizionale", p. 15 nota 33.

connotano la società<sup>192</sup>. In questo caso, il correttivo potrebbe essere individuato non tanto nella *transparency by design*, ma nella *transparency degli input e degli output*<sup>193</sup>.

Questo sistema di controllo dell'output potrebbe anche essere impiegato quando non si è in grado di spiegare come l'algoritmo sia pervenuto a una data decisione ovvero quando l'algoritmo si sia modificato a seguito dell'autoapprendimento<sup>194</sup>. Occorre, pertanto, un controllo non solo a monte, che abbia riguardo alla fase di programmazione, ma anche a valle, cioè che si concentri sull'output, cercando di correggerlo qualora risulti discriminatorio.

Ciò conduce a ritenere sempre necessario il controllo umano sulla decisione algoritmica<sup>195</sup>, la quale non può essere accettata senza che sia sottoposta a revisione.

Nel dibattito europeo sembra essere condivisa l'idea per cui non si può essere sottoposti a decisioni basate unicamente su un trattamento automatizzato. Si tratta di un principio non del tutto innovativo, in quanto già l'art. 22 del GDPR lo contempla. Tale principio può essere ora trasposto a ogni settore in cui si faccia uso di sistemi esperti per assumere una decisione. La prospettiva è stata condivisa anche in Italia, nella specie dal Consiglio di Stato, che ha ripudiato l'idea di decisioni amministrative algoritmiche non seguite da una supervisione umana<sup>196</sup>.

---

<sup>192</sup> A. Chander, *The Racist Algorithm?*, cit. p. 1039.

<sup>193</sup> *Idem*, p. 1039. Cfr. anche *ALGORITHMS AND HUMAN RIGHTS. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications*, Committee of experts on internet intermediaries (MSI-NET), Council of Europe, March 2018, ove, in tema di trasparenza dell'algoritmo, si precisa che: “*there may be a possibility of demanding that key subsets of information about the algorithms be provided to the public, for example which variables are in use, which goals the algorithms are being optimised for, the training data and average values and standard deviations of the results produced, or the amount and type of data being processed by the algorithm*”, p. 38.

<sup>194</sup> A. Chander, *The Racist Algorithm?*, cit., p. 1039.

<sup>195</sup> D. K. Citron, F. Pasquale, *The Scored society*, cit., p. 6, con riferimento al processo di *algorithmic scoring*.

<sup>196</sup> CONSIGLIO DI STATO, SEZ. VI - SENTENZA 13 dicembre 2019, n. 8472, con riferimento agli esiti della procedura nazionale di mobilità attuata con ordinanza ministeriale n. 241/2016, per i posti di docenti della scuola secondaria di primo grado. Censuravano i ricorrenti che la procedura di trasferimento era stata svolta sulla base di un algoritmo non conosciuto e che non avrebbe funzionato correttamente, poiché non aveva tenuto conto delle preferenze espresse pur in presenza di posti disponibili nelle province indicate dai docenti che avevano chiesto il trasferimento, col risultato che i ricorrenti venivano trasferiti in province più lontane da quella di propria residenza o quella comunque indicata come scelta prioritaria. La censura veniva accolta già in sede di primo grado e confermata dal CDS. In particolare, al par. 7.2 della decisione si legge: “*l'impiego di tali strumenti comporta in realtà una serie di scelte e di assunzioni tutt'altro che neutre: l'adozione di modelli predittivi e di criteri in base ai quali i dati sono raccolti, selezionati, sistematizzati, ordinati e messi insieme, la loro interpretazione e la conseguente formulazione di giudizi sono tutte operazioni frutto di precise scelte e di valori, consapevoli o inconsapevoli; da ciò ne consegue che tali strumenti sono chiamati ad operare una serie di scelte, le quali dipendono in gran parte dai criteri utilizzati e dai dati di riferimento utilizzati, in merito ai quali è apparso spesso difficile ottenere la necessaria trasparenza*”. Pur riconoscendo l'importanza della digitalizzazione anche del procedimento amministrativo, evidenziano i giudici che non devono venire meno le garanzie del giusto procedimento. pertanto, due sono

In questo scenario, la sorveglianza umana viene recepita come elemento imprescindibile per realizzare una IA affidabile. Secondo il documento elaborato dal Gruppo di esperti nominato dalla Commissione Europea, il controllo umano può esplicarsi in tre modi: si ha un approccio con intervento umano o *human in the loop* se l'intervento umano è consentito in ogni ciclo decisionale del sistema; l'approccio *human on the loop* prevede il controllo e intervento durante il ciclo di progettazione e il monitoraggio del funzionamento del sistema; infine, l'approccio *human in command* prevede il controllo del sistema di IA nel suo complesso e la capacità di decidere in quali circostanze utilizzare il sistema e in quali no<sup>197</sup>.

La sorveglianza umana consentirebbe di superare alcuni limiti che incontra la trasparenza del funzionamento dell'algoritmo. Tra questi, basti ricordare il limite rappresentato dalla segretezza tutelata dal diritto di proprietà intellettuale<sup>198</sup>, che determina i confini di una *black box* di tipo giuridico e non tecnologico, ma anche la necessità di non rendere il sistema fallace e potenziale oggetto di tentativi di manipolazione o hakeraggio.

## 5. Una conclusione per iniziare

Le questioni giuridiche che emergono sono varie. Dal problema della regolamentazione della produzione di sistemi di IA, chiamato a confrontarsi col diritto di libera iniziativa economica, a quello della trasparenza, che stride con i diritti di proprietà intellettuale tutelati dalla segretezza, fino a giungere alla tematica della responsabilità. Ciò che è dato riscontrare, dinanzi a queste innovazioni tecnologiche, è una 'sete di

---

gli elementi di minima garanzia per ogni ipotesi di utilizzo di algoritmi in sede decisoria pubblica: a) la piena conoscibilità a monte del modulo utilizzato e dei criteri applicati; b) l'imputabilità della decisione all'organo titolare del potere, il quale deve poter svolgere la necessaria verifica di logicità e legittimità della scelta e degli esiti affidati all'algoritmo (par. 12).

Nello stesso senso, si veda la sent. CDS, sez. VI, 4 febbraio 2020, n. 881.

<sup>197</sup> High Level Expert Group, *Orientamenti etici per una IA affidabile*, 8 aprile 2019, p. 18

<sup>198</sup> È l'argomento valorizzato nel noto caso Loomis, *Wisconsin Supreme Court, State v. Loomis*, case 2015AP17-CR, Judgment July 13th 2016, in *130 Harvard L.R.*, 2017, p. 1530 ss. A livello europeo, argomentazioni simili sono state proposte dal Bundesgerichtshof tedesco in una decisione del 28 gennaio 2014. La Corte Suprema tedesca ha rigettato un ricorso volto a conoscere le informazioni alla base dell'algoritmo impiegato da una agenzia di credito, in quanto protetto da segreto. Nondimeno, la Corte ha consentito di conoscere i dati usati per la valutazione di affidabilità circa lo stato di solvibilità del ricevente il prestito. *German Federal Supreme Court (Bundesgerichtshof), Judgment, 28 January 2014 Az. VI ZR 156/13*, la pronuncia è reperibile al link <https://openjur.de/u/677956.html>.

diritto': "ora che l'uomo ha creato un mondo nuovo, non può sottrarsi a ciò che è implicato dall'atto creativo di una nuova realtà: l'ordinarla e il darle delle leggi"<sup>199</sup>.

Questo è il contributo che il giurista può fornire per evitare la "dittatura dell'algoritmo", emblema di una società che annienta l'uomo inteso quale soggetto titolare di diritti, da un lato, e quale decisore, dall'altro<sup>200</sup>.

La visione antropocentrica dello sviluppo dei sistemi di IA, favorita dalle istituzioni europee, dovrebbe spingersi sino a tutelare l'individuo dall'aggressione dei diritti e delle libertà fondamentali da parte delle nuove tecnologie e dovrebbe altresì garantirne la responsabilizzazione.

L'individuo va quindi tutelato, ma anche responsabilizzato. Decisioni importanti rimesse totalmente agli algoritmi, annientando l'intervento o il controllo umano, favoriscono la creazione di un *responsability gap*: vengono messe sotto accusa le macchine e le loro decisioni automatizzate, che non possono difendersi e così diventano un comodo capro espiatorio<sup>201</sup>.

L'obiettivo di questo lavoro è cogliere la "sfida tecnologica" e cercare di trovare delle risposte alla problematica della responsabilità per i danni da produzione e uso di sistemi di IA. L'idea di fondo è che la produzione e l'uso di agenti digitali offrono dei benefici che devono necessariamente essere controbilanciati da forme di responsabilità<sup>202</sup>: *cuius commoda, eius incommoda*.

L'idea proposta da più parti è quella di creare un apparato regolamentare che possa garantire la trasparenza, l'affidabilità della decisione algoritmica. Si tratta di un obiettivo ambizioso: la creazione di un *technological due process*<sup>203</sup> che involga tanto la fase di produzione dell'algoritmo, quanto il momento del suo utilizzo. Ciò significa che le regole si rivolgeranno non solo al produttore, ma anche all'utilizzatore, sia che si tratti di un utilizzatore "comune", sia che si tratti di un soggetto "qualificato" (si pensi alla pubblica amministrazione, agli istituti di credito) che, in quanto tale, si avvale di sistemi di intelligenza artificiale per addivenire a decisioni che incidano su singoli individui.

Solo a valle di questo apparato normativo si collocherà la sanzione, anche penale.

---

<sup>199</sup> S. Cotta, *La sfida tecnologica*, Il Mulino, Bologna, 1968, p. 82.

<sup>200</sup> S. Rodotà, *Il mondo nella rete. Quali i diritti, quali i vincoli*, Laterza, Roma-Bari, VI ed., 2019, p. 37.

<sup>201</sup> *Idem.*, p. 39.

<sup>202</sup> Nello stesso senso, G. Fioriglio, *La "dittatura dell'algoritmo"*, cit., p. 136.

<sup>203</sup> D. K. Citron, *Technological Due Process*, 85 *Wash. U. L. Rev.* 1249, 1260-63 (2008); D. K. Citron e F. Pasquale, *The Scored Society: Due Process for Automated Predictions*, p. 19.



## CAPITOLO II

### INTELLIGENZA ARTIFICIALE E DIRITTO

1. L'Intelligenza Artificiale nell'evoluzione del diritto dell'Unione Europea. L'esigenza di una IA antropocentrica - 1.1 La Risoluzione del Parlamento europeo del febbraio 2017 e l'apertura verso il riconoscimento della personalità giuridica del robot - 1.2. Gli interventi della Commissione europea - 1.2.1 Il Gruppo di Esperti di Alto Livello e gli Orientamenti etici per una IA affidabile - 1.2.2 Il Libro Bianco sull'IA - 1.2.2.1 Individuazione delle aree di intervento della legislazione speciale - 1.3 La definizione di IA - 1.4 La proposta di regolamento del Parlamento europeo e del Consiglio dell'ottobre 2020. L'abbandono dell'idea di una personalità giuridica del robot. L'introduzione della nuova figura dell'*operatore* - 1.5 La proposta di regolamento della Commissione europea - 1.6 Ripartizione degli obblighi tra gli operatori economici coinvolti: dovere di diligenze e posizione di garanzia - 1.7 L'impiego dei sistemi di IA nella giustizia penale - 2. Intelligenza Artificiale e diritto civile - 2.1 La responsabilità del produttore - 2.2 La responsabilità dell'utente - 3. Responsabilità aquiliana e responsabilità penale: fatto illecito e fatto costituente reato - 4. Diritto e tecnologia: regole di convivenza

Il tema della IA accende l'attenzione sul rapporto tra diritto e tecnologia. Per questo, esaminati gli aspetti più tecnici relativi all'IA, bisogna ora domandarsi quale ruolo spetti al diritto, in quanto sistema regolatore delle relazioni sociali e necessario interprete dei relativi cambiamenti<sup>1</sup>.

Il rapporto tra diritto e tecnologia si presenta come intricato e controverso, perché si tratta di due scienze che raramente corrono di pari passo. Solitamente ce n'è una che è sempre più indietro rispetto all'altra, e questa è il diritto. Il diritto insegue la tecnologia e con non poca fatica. Ma, del resto, questa dinamica forse è inevitabile, dato che il diritto, solitamente, regola fenomeni già esistenti e non ne crea di nuovi: prima della regola non può che esistere il fenomeno da regolare. Il binomio, pertanto, non può che essere *tecnologia – diritto*.

Successivamente, una volta individuato il fenomeno, interviene il diritto e a questo punto i termini del binomio possono invertirsi, per divenire *diritto – tecnologia*. Ciò significa che quando il diritto interviene a disciplinare un dato settore, dovrebbe farlo con la capacità di regolare anche aspetti non ancora scoperti o non ancora sviluppati, adoperando cioè una tecnica di codificazione ispirata al principio di neutralità tecnologica. Tale aspetto attiene alla tecnica di normazione, che è necessario segua dei principi fissati dal legislatore. Questo è ciò che si è occupato di fare il legislatore europeo attraverso

---

<sup>1</sup> G. Di Rosa, *Robot e responsabilità per danni*, in *L'algoritmo pensante*, Il Pozzo di Giacobbe, Trapani, 2020, p. 85 ss.

l'elaborazione di principi etici per una IA affidabile, come avremo modo di approfondire appresso.

Quanto all'oggetto della disciplina, si tratta di stabilire delle regole che orientino la produzione dei sistemi di IA e il loro utilizzo.

È stato osservato che nel fare ciò il diritto diviene una "meta-tecnologia"<sup>2</sup>, in quanto mezzo per governare gli strumenti tecnologici attraverso l'individuazione delle condizioni di legittimità per la progettazione, la produzione e l'utilizzo dei sistemi di IA. Così, poiché si tratta di una *tecnica* che disciplina altre *tecniche* e anche il *processo dell'innovazione tecnologica*, la legge può essere considerata una "meta-tecnologia"<sup>3</sup>.

Il tema dell'intelligenza artificiale è stato affrontato nel diritto civile in modo più compiuto rispetto a quanto sinora si è fatto nel diritto penale. La necessità di ripercorrere, seppur brevemente, in questa sede gli sviluppi del tema della responsabilità extracontrattuale per i danni riconducibili a un agente diverso dall'uomo si spiega alla luce della stretta relazione che sussiste tra il *fatto illecito* di cui all'art. 2043 c.c. e il *fatto costituente reato*.

Il maggiore spazio della riflessione civilistica sul punto è stato incentivato anche dalle istituzioni europee. In particolare, nel febbraio 2017 il Parlamento europeo ha approvato una Risoluzione che raccomanda alla Commissione l'elaborazione di una proposta di direttiva in materia di disciplina civilistica dell'intelligenza artificiale applicata alla robotica<sup>4</sup>. Successivamente la Commissione Europea si è espressa sul tema dell'intelligenza artificiale con il *White paper on Artificial Intelligence. A European approach to excellence and trust* e a questa ha poi fatto seguito una proposta di regolamento elaborata dal Parlamento europeo. Emerge, pertanto, come il tema dell'intelligenza artificiale sia affrontato a più livelli.

Allo stato, non si è ancora assistito a un intervento sistematico sul punto. V'è, tuttavia, chi sottolinea che bisognerebbe valorizzare la capacità di mediazione giuridica delle norme esistenti<sup>5</sup>. In particolare, lo sguardo andrebbe rivolto alla disciplina sulla responsabilità da prodotto, la cui matrice comunitaria la rende relativamente uniforme

---

<sup>2</sup> U. Pagallo, *The Laws of Robots. Crimes, Contracts, and Torts*, Springer, Dordrecht, 2013, p. 10.

<sup>3</sup> U. Pagallo, *op. cit.*, p. 11.

<sup>4</sup> Risoluzione recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica, 16 febbraio 2017, 2015/1203INL.

<sup>5</sup> U. Ruffolo, *Per i fondamenti di un diritto della robotica self-learning; dalla machinery produttiva all'auto driverless: verso una "responsabilità da algoritmo"?*, in *Intelligenza artificiale e responsabilità*, a cura di U. Ruffolo, Milano, Giuffrè, 2017, p. 3.

negli Stati membri. Ma non solo. Si assiste anche alla valorizzazione delle norme del codice civile sulla responsabilità extracontrattuale, in particolare l'art. 2050 c.c. sulla responsabilità per l'esercizio di attività pericolosa e, quando ne ricorrono i presupposti, l'art. 2051 c.c. sulla responsabilità per danno da cose in custodia.

Nel prosieguo ci si soffermerà dapprima sugli interventi susseguitisi in ambito europeo, successivamente sul dibattito sorto nella dottrina civilistica italiana sul recupero delle categorie del danno da prodotto difettoso e di alcune norme codicistiche sulla responsabilità aquiliana.

## **1. L'Intelligenza Artificiale nell'evoluzione del diritto dell'Unione Europea. L'esigenza di una IA antropocentrica**

In ambito europeo la reazione dinanzi al fenomeno delle nuove tecnologie è stata piuttosto immediata. Le istituzioni hanno presto compreso l'importanza di interfacciarsi con un settore che offre nuove opportunità ma porta al contempo con sé nuovi rischi<sup>6</sup>.

Certamente l'Unione dispone già di un solido quadro normativo che si presta a disciplinare alcuni aspetti delle nuove tecnologie, le quali non si sono inserite in un contesto privo di normazione<sup>7</sup>. In particolare, la tutela dei dati, sin dalla progettazione, è garantita dal Regolamento Generale sulla Protezione dei Dati (GDPR); quanto ai dati non personali, la loro libera circolazione e il loro trattamento in Europa sono garantiti dal regolamento sulla libera circolazione dei dati<sup>8</sup>; infine, il regolamento sulla cybersicurezza<sup>9</sup> contribuisce a creare un clima di fiducia sulle operazioni che si effettuano *online*.

In questo quadro legislativo, a necessitare di un intervento normativo è soprattutto il settore della responsabilità per danni da sistemi di IA, alla luce delle caratteristiche delle

---

<sup>6</sup> Report from the Expert Group on Liability and New Technologies - New Technologies Formation, *Liability for AI and other emerging technologies*, 2019, p. 32.

<sup>7</sup> *Comunicazione della Commissione Al Parlamento Europeo, Al Consiglio, Al Comitato Economico E Sociale Europeo E Al Comitato Delle Regioni, Creare fiducia nell'intelligenza artificiale antropocentrica*, COM(2019) 168 final, 8.04.2019, par. 2, p. 2.

<sup>8</sup> Regolamento (UE) 2018/1807 del Parlamento europeo e del Consiglio, del 14 novembre 2018, relativo a un quadro applicabile alla libera circolazione dei dati non personali nell'Unione europea.

<sup>9</sup> Regolamento (UE) 2019/881 del Parlamento europeo e del Consiglio, del 17 aprile 2019, relativo all'ENISA, l'Agenzia dell'Unione europea per la cybersicurezza, e alla certificazione della cybersicurezza per le tecnologie dell'informazione e della comunicazione, e che abroga il regolamento (UE) n. 526/2013 («regolamento sulla cybersicurezza»)

nuove tecnologie. Infatti, le applicazioni di IA si presentano come sistemi *complessi, opachi, aperti, autonomi, imprevedibili, in un rapporto di dipendenza con i dati e vulnerabili*<sup>10</sup>.

La *complessità* è dovuta al fatto che si tratta di sistemi altamente sofisticati dal punto di vista tecnologico. Essa è tra le cause dell'*opacità* del sistema, intesa come impossibilità o difficoltà nel comprendere il processo che ha portato alla causazione del danno. Ulteriore aspetto della complessità di tali sistemi è il fatto che siano *sistemi "aperti"*. L'espressione allude all'interazione del sistema con altri sistemi o con dati presi dalla rete, aspetto che rende l'applicazione *vulnerabile*, in quanto esposta ai rischi di attacchi dall'esterno. Inoltre, spesso le applicazioni di IA sono *autonome*, poiché agiscono senza un controllo o una supervisione umana; ciò rende il loro *comportamento non sempre prevedibile*.

Queste caratteristiche sono ancor più accentuate nei sistemi di apprendimento automatico. Come detto, la macchina assume la decisione in modo autonomo, ma le vengono in origine fornite delle conoscenze. Vi sono degli *input* inseriti dall'uomo, attraverso i quali la macchina potrà elaborare i dati percepiti dall'esterno e giungere poi a una decisione. Nel momento in cui l'uomo programma gli *input* iniziali della macchina non sarà in grado di conoscere quali scelte il sistema assumerà nelle varie esperienze in cui si troverà coinvolta. Questo è il primo livello di incertezza, che si traduce in una imprevedibilità *ex ante* delle decisioni della macchina. Talvolta, tuttavia, l'incertezza si proietterà anche nel momento successivo alla decisione. Si tratta dei casi in cui sarà impossibile capire in che modo la macchina sia giunta ad assumere una certa decisione. È il c.d. *black box problem*, che determina una opacità *ex post* della decisione, che accompagna la sua imprevedibilità *ex ante*.

Queste peculiarità aprono nuovi scenari in punto di sistemi di responsabilità per i danni causati dai sistemi di IA, dinanzi ai quali il legislatore europeo non poteva certo rimanere silente.

Gli interventi sono da attribuire tanto al Parlamento Europeo, quanto alla Commissione, come di seguito analizzeremo.

---

<sup>10</sup> *Report from the Expert Group on Liability and New Technologies*, cit, p. 32 ss, ma anche COM (2020) 64 final del 19.02.2020, che riprende proprio il documento del Gruppo di esperti

## 1.1 La Risoluzione del Parlamento europeo del febbraio 2017 e l'apertura verso il riconoscimento della personalità giuridica del robot

La Risoluzione del Parlamento europeo del 16 febbraio 2017, recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)), rappresenta il primo atto politico che affronta in modo unitario i temi connessi alla robotica prendendo in considerazione le conseguenze anche giuridiche della diffusione della IA<sup>11</sup>. La scelta di anticipare i singoli Parlamenti nazionali non è di poco momento, poiché alla base vi è l'auspicio di una normativa di settore armonizzata e il più possibile simile all'interno di ciascuno Stato membro<sup>12</sup>.

La Risoluzione offre un quadro sulle “norme di diritto civile sulla robotica”. Essa fissa dei principi generali e affronta, tra gli altri, il tema della responsabilità.

Va anzitutto effettuata una notazione terminologica: a dispetto della nomenclatura adoperata nell'intitolazione della Risoluzione, il testo si rivolge anche all'intelligenza artificiale, considerando i termini “robot” e “IA” come intercambiabili<sup>13</sup>. Ciò conferma l'impostazione metodologica che tende ad equiparare la robotica e l'IA, pur trattandosi di settori diversi, che certamente si intersecano in alcuni momenti, come si è già avuto modo di notare nel cap. 1, par. 1, cui si rinvia.

La Risoluzione prende le mosse dal constatato sviluppo della ricerca tecnologica che ha portato alla produzione di robot dotati di una certa autonomia, circostanza, quest'ultima, che rende essenziale la questione della responsabilità giuridica derivante dall'azione dannosa di un robot.

Per “autonomia” la Risoluzione intende “la capacità di prendere decisioni e metterle in atto nel mondo esterno, indipendentemente da un controllo o un'influenza esterna”<sup>14</sup>. Da tale autonomia discende che il robot non può essere considerato un “mer[o] strument[o] nelle mani di altr[o] attor[e]” e ciò impone di interrogarsi sotto un duplice fronte. Da un canto, ci si dovrà porre il problema dell'eshaustività del quadro normativo attualmente esistente ovvero della necessità di nuovi principi o regole che definiscano la

---

<sup>11</sup> A. Amidei, *Intelligenza artificiale e diritto*, cit.

<sup>12</sup> M. Bassini, L. Liguori, O. Pollicino, *Sistemi di intelligenza artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, in *Intelligenza artificiale, protezione dei dati personali e regolazione*, a cura di F. Pizzetti, 2018, p. 338.

<sup>13</sup> M. Bassini, L. Liguori, O. Pollicino, *Sistemi di intelligenza artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, cit., p. 336.

<sup>14</sup> lett. AA.

responsabilità legale dei vari attori per le azioni e le omissioni imputabili ai robot. Dall'altro, la Relazione pone il quesito del possibile riconoscimento di una personalità giuridica dei robot, posto che, nell'attuale quadro giuridico, i robot non possono essere considerati responsabili in proprio per i danni cagionati a terzi.

Il Parlamento adotta allora un atteggiamento duplice. Da un lato, osserva che le norme sulla responsabilità per i danni da prodotto e, più in generale, le norme sulla responsabilità civile sono sufficienti a garantire un'allocazione dei danni allorché il robot non sia dotato di autonomia. Dall'altro, tuttavia, è consapevole del fatto che, nel caso di azioni imputabili a un robot in grado di prendere decisioni autonome, le norme tradizionali in tema di responsabilità risultano insufficienti, poiché coprono i casi in cui la causa di un'azione o di un'omissione del robot sia riconducibile a uno specifico agente umano (il fabbricante, l'operatore, il proprietario o l'utilizzatore) e nei quali l'agente avrebbe potuto prevedere e evitare il comportamento dannoso del robot.

Così, la Risoluzione riconosce la necessità di procedere all'integrazione del quadro normativo vigente, con particolare riferimento ai casi di danni causati dalla nuova generazione di robot dotati di capacità di adattamento e di apprendimento, implicanti un certo grado di imprevedibilità nel loro comportamento, attesa la loro capacità di *deep learning*. Ciononostante, osserva, il Parlamento, che “almeno nella fase attuale, la responsabilità deve essere imputata a un essere umano e non a un robot”<sup>15</sup>. In tale contesto, il Parlamento non manca di evidenziare la necessità di istituire “uno status giuridico specifico per i robot nel lungo termine, di modo che almeno i robot autonomi più sofisticati possano essere considerati come persone elettroniche responsabili di risarcire qualsiasi danno da loro causato”. Ma non solo.

Si dà spazio anche a un eventuale riconoscimento della personalità elettronica dei robot che prendono decisioni autonome o che interagiscono in modo indipendente con soggetti terzi<sup>16</sup>.

La proposta era già stata sostenuta da una parte della dottrina, anche penalistica<sup>17</sup>, che si era occupata del tema. Su questo punto, tuttavia, è bene evidenziare che in senso contrario al riconoscimento di forme di personalità giuridica dei robot o dei sistemi di IA si è pronunciato il Comitato economico e sociale europeo (EESC). La

---

<sup>15</sup> Punto 56.

<sup>16</sup> Punto 59, lett. f)

<sup>17</sup> In particolare, per una responsabilità penale personale del robot, G. Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, Berlino, 2015.

deresponsabilizzazione dell'uomo porterebbe, secondo la EESC, a svuotare il sistema della responsabilità della sua funzione di prevenzione dei danni, oltre che di quella riparatrice, così favorendo fenomeni di abuso da parte degli utilizzatori e dei produttori<sup>18</sup>. Le stesse perplessità sono state sollevate dal Gruppo di Esperti su Responsabilità e nuove tecnologie<sup>19</sup>, che ha criticato la Posizione assunta dal Parlamento Europeo<sup>20</sup>, considerato che i danni causati da un sistema di IA possono essere eziologicamente imputati a persone fisiche o giuridiche e, ove ciò non sia possibile, nuovi interventi legislativi dovranno occuparsi della questione, senza però creare nuove categorie di soggetti giuridici<sup>21</sup>.

Il Parlamento sembra aver ascoltato queste critiche; e difatti l'idea della personalità giuridica del robot non è stata riproposta nella più recente Risoluzione del 20 ottobre 2020 (2020/2014 (INL)), sempre a firma del Parlamento, con la quale è stata formulata una proposta di regolamento del Parlamento europeo e del Consiglio sulla responsabilità per il funzionamento dei sistemi di intelligenza artificiale, di cui si dirà nei paragrafi a seguire.

Ai fini dell'imputazione della responsabilità civile, il Parlamento individua due possibili approcci tra i quali la Commissione potrà scegliere nell'adozione del futuro strumento legislativo: si tratta dell'approccio della responsabilità oggettiva e di quello della gestione dei rischi.

Come noto, la responsabilità oggettiva “richiede una semplice prova del danno avvenuto e l'individuazione di un nesso di causalità tra il funzionamento lesivo del robot e il danno subito dalla parte lesa”. Per contro, “l'approccio di gestione dei rischi non si concentra sulla persona ‘che ha agito con negligenza’ in quanto responsabile a livello individuale, bensì sulla persona che, in determinate circostanze, è in grado di minimizzare i rischi e affrontare l'impatto negativo”<sup>22</sup>.

---

<sup>18</sup> Comitato economico e sociale europeo, *Artificial Intelligence for Europe*, 2019, p. 1. Il file in formato pdf è disponibile al sito: <https://www.eesc.europa.eu/it/our-work/publications-other-work/publications/artificial-intelligence-europe> (ultima consultazione agosto 2020).

<sup>19</sup> Nella Comunicazione COM (2020) 64 final del 19.02.2020, p. 2 nota 11, versione inglese, si legge che il Gruppo, istituito dalla Commissione, che raccoglie esperti in tema di responsabilità civile per i danni da prodotto difettoso, cui è stato affidato il compito di redigere dei principi guida circa l'adattamento del quadro legislativo già esistente alle nuove tecnologie.

<sup>20</sup> Report from the Expert Group on Liability and New Technologies, *cit.*, pp. 37-38.

<sup>21</sup> *Idem*, in particolare p. 38.

<sup>22</sup> Punti 54 e 55

Questa alternatività del sistema di regolamentazione della responsabilità è stata criticata in dottrina<sup>23</sup>, la quale non ha mancato di notare che i due modelli ben potrebbero integrarsi nell'ambito della disciplina della responsabilità del produttore<sup>24</sup>, costituendo “diversi momenti del medesimo fenomeno”, al fine di garantire un'effettiva allocazione delle responsabilità tra i soggetti coinvolti. Così è stato proposto di ricorrere al regime di *strict liability* per l'accertamento dell'imputazione della responsabilità, mentre il *risk management approach* dovrebbe essere utilizzato ai fini dell'identificazione del soggetto responsabile. Di conseguenza, sarebbe sufficiente il solo accertamento del nesso causale tra danno e azione del robot, mentre il soggetto cui attribuire la responsabilità oggettiva dovrebbe essere individuato nella persona che, tra i soggetti della catena produttiva, sia reputata “più vicin[a] al prodotto” alla luce del tipo di malfunzionamento e dunque più idonea a gestirne a monte il relativo rischio<sup>25</sup>.

L'idea della responsabilità oggettiva è favorevolmente considerata dalla dottrina italiana la quale osserva che, a ritenere diversamente, vi sarebbe spazio per preconstituire forme di esclusione della responsabilità che rischiano di scaricare i costi del danno sul singolo o sulla collettività<sup>26</sup>. In questo senso, va tenuto fermo il principio per cui dietro le macchine deve sempre essere rintracciabile l'elemento umano, almeno fintantoché le macchine agiscano in modo “non determinato dalle caratteristiche del programma che le muove”<sup>27</sup>.

Nondimeno, la responsabilità oggettiva comporta un aggravio in termini di costi a carico del produttore; la questione è presa in considerazione dal Parlamento, il quale, per temperare le conseguenze economiche del regime di responsabilità oggettiva, suggerisce di istituire un regime assicurativo obbligatorio per i soggetti di cui si compone la catena produttiva, così configurando uno schema risarcitorio simile a quello già in atto per i danni da sinistro stradale. Ma oltre a ciò, il Parlamento propone la creazione di un fondo per garantire il risarcimento quando il danno causato dal robot non è assicurato<sup>28</sup>, dalla cui costituzione far dipendere il beneficio di una responsabilità limitata per il produttore, il programmatore, il proprietario o l'utente. Il robot dovrebbe essere associato al singolo

---

<sup>23</sup> A. Amidei, *Robotica intelligente e responsabilità: profili e prospettive evolutive del quadro normativo europeo*, in *Intelligenza artificiale*, cit. p. 86.

<sup>24</sup> G. Capilli, *Responsabilità e robot*, in *Nuova giur. civ.*, 2019, 3, 621.

<sup>25</sup> *Cit.*, p. 87.

<sup>26</sup> R. Bifulco, *Intelligenza artificiale, internet e ordine spontaneo*, in F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018, p. 390.

<sup>27</sup> *Ibidem*, p. 390.

<sup>28</sup> Punto 59, lett. b).

fondo attraverso un numero di immatricolazione individuale; in tal modo, coloro che interagiscono col robot possono acquisire le informazioni sulla natura del fondo, sui limiti della responsabilità in caso di danni alle cose, sui nomi e sulle funzioni dei contributori.

Il Parlamento conclude chiedendo alla Commissione, a norma dell'articolo 225 TFUE, di presentare una proposta di direttiva sulle norme di diritto civile sulla robotica, sulla base dell'articolo 114 TFUE.

Un ulteriore invito rivolto alla Commissione affinché intervenga sul tema della IA proviene dal Consiglio europeo nell'ottobre del 2017. L'obiettivo è garantire che lo sviluppo di tali sistemi sia accompagnato da “un alto livello di protezione dei dati personali, dei diritti digitali e di standard etici”<sup>29</sup>.

## 1.2. Gli interventi della Commissione europea

Alla Risoluzione del Parlamento Europeo hanno fatto seguito diverse comunicazioni della Commissione Europea<sup>30</sup>, tutte tese a sottolineare la necessità che lo sviluppo della IA avvenga in un quadro giuridico e etico adeguato, nel rispetto dei valori europei e dei diritti sanciti nella Carta dei diritti fondamentali dell'Unione Europea<sup>31</sup>, e ciò al fine di creare una IA antropocentrica e affidabile<sup>32</sup>.

La Commissione si è anche interessata del tema della responsabilità. La questione è stata oggetto di dibattito da parte dell'Expert Group on Liability and New Technologies, istituito dalla Commissione, che nel 2019 ha pubblicato il documento dal titolo “*Liability for Artificial Intelligence and other Emerging Digital Technologies*”<sup>33</sup>, dando atto delle sfide poste dalle nuove tecnologie in tema di gestione dei rischi e di responsabilità.

Successivamente, il 19 febbraio 2020 è stato pubblicato il Libro Bianco della Commissione Europea sull'Intelligenza artificiale, dal titolo “*An european approach to excellence and trust*”.

---

<sup>29</sup> EUCO 14/17, p. 7, disponibile al sito del Consiglio europeo <http://data.consilium.europa.eu/doc/document/ST-14-2017-INIT/en/pdf>

<sup>30</sup> Ad esempio, *Artificial Intelligence for Europe*, COM (2018) 237 final del 25 aprile 2018; *Coordinated Plan on AI*, COM (2018) 795 final del 7 dicembre 2018; COM(2019) 168 final dell'8 aprile 2019, tutte disponibili al sito della Commissione europea. In generale, va notato che gli interventi della Commissione Europea in tema di IA sono stati plurimi e di vario tipo, pertanto ci soffermeremo solo su alcuni di essi.

<sup>31</sup> COM (2018) 237 final, p. 4.

<sup>32</sup> COM(2019) 168 final, par. 2, p. 2 ss.

<sup>33</sup> <https://op.europa.eu/it/publication-detail/-/publication/1c5e30be-1197-11ea-8c1f-01aa75ed71a1/language-en>

### 1.2.1 Il Gruppo di Esperti di Alto Livello e gli Orientamenti etici per una IA affidabile

Sin dal momento in cui la Commissione ha iniziato a occuparsi del tema dell'IA, essa ha chiarito che l'approccio prescelto sarebbe stato "antropocentrico": in quest'ottica, l'IA non deve essere fine a se stessa, ma deve essere intesa come uno strumento volto al raggiungimento del benessere per l'uomo<sup>34</sup>. L'idea di fondo è che la ricerca in tema di IA debba conformarsi, oltre che alla legge, anche a principi etici<sup>35</sup> e ciò comporta che anche la legge che regola il campo dell'IA sia ispirata a tali principi.

Tale esigenza aveva indotto la Commissione a istituire, nel giugno del 2018, un Gruppo di Esperti ad alto livello sull'intelligenza artificiale. Il risultato del lavoro di tale gruppo è il documento intitolato "Orientamenti etici per un'IA affidabile"<sup>36</sup>.

Il documento valorizza l'idea di un sistema di sviluppo della IA antropocentrico<sup>37</sup>, che garantisca il rispetto dei diritti fondamentali sanciti, tra l'altro, nei trattati UE e nella Carta di Nizza<sup>38</sup>. Vengono, inoltre, individuate le tre componenti che un'IA deve possedere affinché possa ritenersi "affidabile". Esse sono: a) legalità; b) eticità; c) robustezza<sup>39</sup>. La *legalità* richiede che l'IA rispetti tutte le leggi e i regolamenti applicabili;

---

<sup>34</sup> Tale approccio è esplicitato nella comunicazione della Commissione indirizzata al Parlamento Europeo, alla Commissione europea su economica e sociale, nonché alla Commissione delle Regioni, ove si legge: *"The European AI strategy and the coordinated plan make clear that trust is a prerequisite to ensure a human-centric approach to AI: AI is not an end in itself, but a tool that has to serve people with the ultimate aim of increasing human well-being"*. COM (2019) 168, p. 1, disponibile su <https://ec.europa.eu/digital-single-market/en/news/communication-building-trust-human-centric-artificial-intelligence>

<sup>35</sup> COM (2019) 168, p. 2

<sup>36</sup> Documento reso pubblico l'8 aprile 2019. Una prima versione di tale documento, pubblicata il 18 dicembre 2018, è stata sottoposta a una consultazione pubblica durante la quale oltre 500 partecipanti hanno fornito un loro riscontro. Di tali osservazioni si è tenuto conto ai fini dell'elaborazione della versione finale del testo.

<sup>37</sup> Nella parte finale delle Linee guida etiche, si fornisce una definizione di "IA antropocentrica", ove si legge: *"L'approccio antropocentrico all'IA è volto a garantire che i valori umani rivestano un ruolo centrale nelle modalità di sviluppo, distribuzione, utilizzo e monitoraggio dei sistemi di IA, garantendo il rispetto dei diritti fondamentali, tra cui quelli sanciti nei trattati dell'Unione europea e nella Carta dei diritti fondamentali dell'Unione europea, accomunati dal riferimento a un fondamento condiviso radicato nel rispetto della dignità umana, nei quali l'essere umano gode di uno status morale unico e inalienabile. Ciò implica anche il rispetto dell'ambiente naturale e di altri esseri viventi che fanno parte dell'ecosistema umano e un approccio sostenibile che consenta alle generazioni future di prosperare"*.

<sup>38</sup> Una ricerca del Consiglio d'Europa mostra come l'IA possa avere un forte impatto su un largo numero di diritti fondamentali, <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>.

<sup>39</sup> Pag. 6, par. (15). Nella sintesi collocata in apertura del documento, si legge quanto segue: *Un'IA affidabile si basa su tre componenti che dovrebbero essere presenti durante l'intero ciclo di vita del sistema: a) legalità, l'IA deve ottemperare a tutte le leggi e ai regolamenti applicabili, b) eticità, l'IA deve assicurare l'adesione a principi e valori etici, e c) robustezza, dal punto di vista tecnico e sociale poiché, anche con le migliori intenzioni, i sistemi di IA possono causare danni non intenzionali. Ciascuna componente in sé*

il requisito dell'*eticità* è soddisfatto allorché sia assicurata l'adesione a principi e valori etici; infine, la *robustezza* è da intendersi dal punto di vista tecnico, ma anche sociale, in quanto bisogna tenere in debita considerazione il contesto e l'ambiente in cui il sistema opera<sup>40</sup>.

Quanto all'aspetto della "legalità" dell'IA, esso non è molto approfondito dal Gruppo di esperti, che si limita a osservare che un corpus normativo di riferimento è già presente a livello europeo. A parere del Gruppo di esperti, il diritto primario e derivato dell'UE, seppur non pensato in origine per disciplinare sistemi di IA, si mostra "pertinente" per lo sviluppo, la distribuzione e l'utilizzo di tali sistemi<sup>41</sup>.

Gli orientamenti offrono poi indicazioni per promuovere e garantire la seconda e terza componente di un'IA affidabile, cioè l'*eticità* e la *robustezza*<sup>42</sup>.

Il documento delinea le basi di una IA affidabile secondo un approccio basato sui diritti fondamentali, pertanto identifica e descrive i principi che garantiscono l'*eticità* e la *robustezza* della IA. I principi sono poi declinati in sette requisiti che i sistemi di IA dovrebbero attuare e soddisfare nel ciclo della loro vita: in questo modo il Gruppo di Esperti offre gli spunti per la *realizzazione concreta* di una IA affidabile. Infine, sempre nell'ottica della concretezza, è elaborata una lista di controllo per la valutazione dell'affidabilità dell'IA, al fine di rendere operativi i sette requisiti individuati.

Il ragionamento per la costruzione di una IA affidabile prende le mosse da quattro principi etici, desunti o derivati dai diritti fondamentali previsti dal diritto internazionale, dai trattati dell'UE e dalla Carta di Nizza. Si tratta del principio del rispetto dell'autonomia umana, della prevenzione dei danni, dell'equità e dell'esplicabilità.

In ordine al primo principio, si precisa che gli esseri umani che interagiscono coi sistemi di IA devono poter mantenere la loro libertà di autodeterminazione, non vedendo

---

*è necessaria ma non sufficiente per realizzare un'IA affidabile. Idealmente le tre componenti operano armonicamente e si sovrappongono; qualora, nella pratica, si dovessero verificare tensioni tra di esse la società dovrebbe adoperarsi per risolverle", p. 2.*

<sup>40</sup> Par. (27), pag. 8.

<sup>41</sup> Pag. 7, par. (22) del documento: "Le fonti giuridiche pertinenti sono, a titolo esemplificativo, il diritto primario dell'UE (i trattati dell'Unione europea e la sua Carta dei diritti fondamentali), il diritto derivato dell'UE (ad esempio il regolamento generale sulla protezione dei dati, le direttive antidiscriminazione, la direttiva macchine, la direttiva sulla responsabilità dei prodotti, il regolamento sulla libera circolazione dei dati non personali, il diritto dei consumatori e le direttive in materia di salute e sicurezza sul lavoro), ma anche i trattati ONU sui diritti umani e le convenzioni del Consiglio d'Europa (come la Convenzione europea dei diritti dell'uomo) e numerose leggi degli Stati membri dell'UE. Oltre alle norme applicabili orizzontalmente, esistono varie norme specifiche per settore applicabili a particolari applicazioni di IA (ad esempio il regolamento sui dispositivi medici nel settore sanitario)".

<sup>42</sup> Par. (24)

manipolate le proprie scelte. Al contempo, i sistemi di IA devono rispettare il principio di prevenzione dei danni<sup>43</sup> e, a tal fine, devono essere tecnicamente robusti. Il terzo principio è quello di equità<sup>44</sup>, che conosce una dimensione sostanziale e una procedurale: sotto il primo profilo, è necessario garantire un’equa distribuzione dei costi e dei benefici e evitare situazioni discriminatorie; sotto il secondo profilo, le decisioni elaborate dai sistemi di IA e dagli esseri umani che li utilizzano devono poter essere revisionate. Infine, il principio dell’esplicabilità<sup>45</sup> è attuato in presenza di processi trasparenti di funzionamento dei sistemi teso a rendere spiegabili e comprensibili le decisioni, anche attraverso l’indicazione degli scopi e delle capacità del sistema.

Una volta individuati i principi, essi vengono trasposti in sette requisiti che i sistemi devono possedere e che i produttori devono rispettare per realizzare una IA affidabile: intervento e sorveglianza umani; robustezza tecnica e sicurezza; riservatezza e governance dei dati; trasparenza; diversità, non discriminazione e equità; benessere sociale e rispetto dell’ambiente; accountability, cioè forme di responsabilità.

Nell’ottica di fornire indicazioni concrete per la creazione di una IA affidabile, il documento individua dei metodi di realizzazione tecnici e non tecnici. Si tratta di un insieme di indicazioni utili a orientare la regolamentazione della fase di produzione, posto che i produttori e gli sviluppatori sono individuati come i primi soggetti tenuti al rispetto delle regole per la realizzazione di sistemi affidabili, chiamati a garantire un’etica *by design*<sup>46</sup>.

I metodi non tecnici concernono le modalità di regolamentazione della produzione e dell’utilizzo della IA. Interessante è la proposta di approcci di regolamentazione non solo pubblica, ma anche privata. Già a livello di organizzazione d’impresa si suggerisce l’adozione di codici di condotta<sup>47</sup> e codici deontologici professionali. Accanto a questa regolamentazione privata, dovrebbe trovare sviluppo la regolamentazione pubblica, che non deve ostacolare la produzione e la libertà d’impresa, ma favorirla. Per tale ragione, viene indicata come utile anche la previsione di sistemi di accreditamento, sull’esempio delle norme ISO, al fine di prevedere una sorta di etichetta “IA affidabile” che, facendo riferimento a norme tecniche specifiche, confermi che il sistema risponde ai requisiti di

---

<sup>43</sup> Par. 51.

<sup>44</sup> Par. 52.

<sup>45</sup> Par. 53.

<sup>46</sup> Par. 98.

<sup>47</sup> Par. 105.

sicurezza, robustezza tecnica e spiegabilità<sup>48</sup>, pur precisando che i meccanismi di certificazione non sostituiranno la responsabilità. Lo sforzo è quello di trovare un bilanciamento tra promozione dello sviluppo economico e produttivo, da un lato, e tutela degli utenti finali, dall'altro.

Il documento si conclude con la descrizione di una “versione pilota” della lista di controllo<sup>49</sup> da sottoporre alle aziende per la valutazione dell'affidabilità della IA, con la proposta di inserire, all'interno delle imprese, una struttura di governance per verificarne l'attuazione.

Il documento, pubblicato nel 2018, presenta degli spunti interessanti. Ha il merito di inquadrare le problematiche relative alla IA sotto il profilo tecnico e normativo. La sfida posta dalla IA, del resto, è proprio quella di instaurare un nuovo dialogo tra tecnica e diritto: il diritto non può regolare lo sviluppo tecnologico sino ad annullarlo, mentre la tecnica deve svilupparsi e crescere nel rispetto delle norme di diritto.

Un ulteriore aspetto merita, a mio avviso, la giusta attenzione: la regolamentazione, in un settore così altamente sperimentale e innovativo, non può prescindere dal contributo dei soggetti interessati; solo attraverso l'incontro tra i soggetti pubblici e privati il diritto comprenderà la tecnica e potrà regolamentarla senza ostacolarla, garantendo, al contempo, il rispetto dei principi fondamentali posti a tutela della persona e della dignità umana. Giustamente, l'interlocutore privato è individuato nelle imprese, le quali sono viste come destinatarie della normazione di settore.

### **1.2.2 Il Libro Bianco sull'IA**

Su tale scia, nel White Paper del febbraio 2020 si osserva che gli sviluppatori di sistemi di IA sono già tenuti a seguire le norme europee in tema di diritti fondamentali, protezione dei consumatori, sicurezza dei prodotti e responsabilità. Il quesito che si pone il legislatore europeo è relativo alla necessità di implementare il corpus normativo già esistente introducendo nuove regole<sup>50</sup>. L'opinione è espressa in tal senso.

Secondo la Commissione, il quadro legislativo esistente dovrà essere migliorato per far fronte ai nuovi rischi connessi all'IA. L'innovazione legislativa, dipoi, è resa

---

<sup>48</sup> Par. 106.

<sup>49</sup> Par. 113.

<sup>50</sup> White Paper, cit., p. 10.

necessaria al fine di evitare una frammentazione a livello nazionale, che avrebbe un impatto notevole sul mercato unico europeo. Regole diverse per ciascun Stato membro scoraggerebbero le imprese che vogliono investire sull'IA; per contro, un quadro normativo chiaro a livello europeo ne faciliterebbe l'accesso nel mercato europeo e le renderebbe competitive a livello globale<sup>51</sup>.

Nel documento si sottolinea anche la necessità di implementare la legislazione relativa alla responsabilità.

Su questo punto, taluni dubbi sono evidenziati con riguardo alla disciplina della responsabilità da prodotto difettoso. Si nota, ad esempio, che tale disciplina si applica ai "prodotti", ma non anche ai servizi. Resterebbero fuori dal suo ambito di applicazione i servizi basati sulla tecnologia intelligente, quali servizi sanitari, finanziari e di trasporto. Ancora, nei sistemi di machine learning taluni rischi possono emergere solo dopo che essi siano collocati sul mercato e non essere prevedibili prima, mentre nella disciplina attualmente esistente l'allocazione della responsabilità ha riguardo ai difetti presenti al momento in cui il prodotto è immesso nel mercato. Una questione da affrontare è, pertanto, il tema dell'allocazione della responsabilità tra i diversi operatori economici nella catena produttiva.

Si riconosce che nuove norme sono necessarie per coprire i nuovi rischi connessi alle nuove tecnologie, affinché chi è danneggiato dall'impiego di sistemi di IA riceva una tutela eguale a chi è danneggiato da altri sistemi tecnologici. Per questa ragione, la conclusione cui giunge la Commissione è nel senso di un intervento legislativo duplice: da un lato, occorrerà apportare modifiche alla legislazione attualmente esistente; dall'altro, un nuovo impianto normativo dedicato alla sola IA potrebbe rendersi necessario al fine di uniformare le legislazioni statali e far fronte alle nuove sfide poste dallo sviluppo tecnologico<sup>52</sup>.

La nuova regolamentazione, tuttavia, non dovrà paralizzare lo sviluppo tecnologico, pertanto lo sforzo del legislatore consiste nel trovare un punto di equilibrio tra libertà di iniziativa economica e tutela dei diritti fondamentali della persona. In tale ottica, la Commissione sostiene una regolamentazione che segua un approccio basato sul rischio, al fine di garantire la proporzionalità dell'intervento regolatore<sup>53</sup>.

---

<sup>51</sup> White Paper, cit., p. 15.

<sup>52</sup> White Paper, p. 16.

<sup>53</sup> White Paper, par. 5, let. C, pag. 19.

Per raggiungere tale obiettivo, la Commissione propone una differenziazione sul piano della disciplina, distinguendo tra applicazioni di IA che comportano un alto rischio e applicazioni a basso rischio, con la conseguenza che solo per le prime sarebbe imposta l'osservanza di talune norme e si sarebbe tenuti a dimostrare il possesso di determinati requisiti; mentre, per le seconde la nuova normativa si applicherebbe su adesione volontaria.

La distinzione tra IA ad alto rischio e a basso rischio presuppone l'elaborazione di un criterio discrezionale tra rischi. A tal fine, sono proposti due criteri cumulativi per la classificazione come "applicazione di IA a alto rischio", mentre la categoria di "IA a basso rischio" sarebbe individuata in modo residuale.

Si legge, nel Libro bianco, che anzitutto è tale l'applicazione di IA impiegata in settori in cui, date le caratteristiche delle attività abitualmente svolte, è possibile prevedere dei rischi significativi. In questo modo, l'intervento normativo sarebbe indirizzato ai soli settori in cui i rischi "sono generalmente ritenuti più probabili". I settori dovranno essere tassativamente indicati nella nuova normativa, con possibilità di sottoporre la lista a degli aggiornamenti. Così sarebbe soddisfatta l'esigenza di certezza del diritto anche in settori emergenti.

Ma che l'attività rientri tra quelle indicate nella lista non è condizione sufficiente. Occorre, altresì, che l'applicazione di IA appartenente a uno dei settori indicati sia "utilizzata in modo tale da poter generare rischi significativi". Questo secondo criterio funge da ulteriore filtro all'interno dei settori individuati alla luce del primo criterio: "non tutti gli usi dell'IA nei settori selezionati comportano necessariamente rischi significativi"<sup>54</sup>. Pertanto, individuato un settore ad alto rischio, bisogna avere riguardo al tipo di *uso* che si faccia delle applicazioni di IA che operano all'interno dello stesso: la valutazione del livello di rischio dipende dall'utilizzo del sistema. Un impiego sicuramente idoneo a generare rischi significativi è quello che produce effetti sui diritti di un individuo o della società, così come gli usi che presentano il rischio di lesioni, morte o altri danni materiali o immateriali significativi. Si suggerisce, allora, che una selezione degli impieghi della IA da cui derivi un rischio significativo potrebbe essere fatta avuto riguardo all'impatto che tale tecnologia ha sui soggetti interessati.

---

<sup>54</sup> Il passaggio è chiarito con un esempio: "*Ad esempio, per quanto l'assistenza sanitaria in genere possa essere certamente un settore rilevante, un eventuale difetto del sistema di prenotazione degli appuntamenti in un ospedale non presenta, in linea di massima, rischi tali da giustificare un intervento legislativo*", p. 19-20.

Dopo aver enunciato tale regola, la Commissione ammette delle eccezioni. Osserva, in particolare, che in determinati casi, certe attività possono essere sottoposte alle regole speciali sebbene non appartengano ad alcun settore considerato di alto rischio, proprio in ragione dell'uso che ne viene fatto. Pertanto, se, in generale, la disciplina speciale è rivolta a fattispecie individuate attraverso l'applicazione di un criterio generale prima, cioè l'appartenenza a un settore ad alto rischio, e particolare poi, cioè l'uso che ingeneri rischi significativi, in altri casi ben può procedersi all'individuazione di attività senza passare dal criterio generale di cui si è detto<sup>55</sup>.

Il richiamo al concetto di rischio evoca il ragionamento sotteso alla formulazione di regole cautelari. E, in effetti, questo è l'approccio seguito dalla Commissione all'interno del Libro bianco. La Commissione non si preoccupa di stabilire le regole del piano sanzionatorio, ma interviene su un livello cronologicamente precedente, cioè quello della prevenzione del rischio. A tal fine, sono individuate le macro categorie su cui la futura legislazione dovrà intervenire introducendo prescrizioni preventive, finalizzate a scongiurare la concretizzazione dei rischi connessi all'uso delle nuove tecnologie di IA.

### **1.2.2.1 Individuazione delle aree di intervento della legislazione speciale**

Una volta individuati i soggetti destinatari della legislazione speciale volta a prevenire i rischi da uso di sistemi di IA, la Commissione precisa che le prescrizioni di tale legislazione dovranno concernere cinque elementi essenziali, che sono: i dati di addestramento; la tenuta di dati e registri; le informazioni da fornire; la robustezza e la precisione; la supervisione umana; specifiche prescrizioni per particolari sistemi di IA, come quelle per l'identificazione biometrica remota. Il risultato è l'individuazione di specifiche aree di intervento, per conseguire determinati risultati ispirati alle esigenze di garanzia dei diritti fondamentali degli individui e di sicurezza.

Quanto al primo punto, i dati di addestramento<sup>56</sup>, ci si prefigge di intervenire già nel primo anello di elaborazione di un sistema di IA. Si tratta del momento in cui vengono

---

<sup>55</sup> Stando a quanto si legge nel Libro bianco, sarebbe da considerare attività di alto rischio l'uso delle applicazioni di IA nei processi di selezione del personale e nelle situazioni che incidono sui diritti dei lavoratori, nonché l'uso di applicazioni di IA ai fini di identificazione biometrica remota e l'impiego di altre tecnologie di sorveglianza intrusive, p. 20.

<sup>56</sup> Par. D), lett. a), p. 21.

inseriti i dati nel sistema, che costituiranno *l'input* che orienterà la futura azione del sistema di IA. Come è già stato evidenziato nel cap. 1, i dati rivestono un ruolo cruciale, in quanto da essi dipendono le conoscenze della macchina e le sue future azioni. Per tale ragione, questa prima fase è di fondamentale importanza ed è stato giustamente individuato dalla Commissione come primo momento della catena produttiva in cui il legislatore dovrà intervenire.

In particolare, il rischio da scongiurare attiene alle possibili informazioni parziali o incomplete che siano inserite nel sistema di IA, dalle quali dipenda la formulazione di un risultato discriminatorio da parte del sistema. Come già notato<sup>57</sup>, la discriminazione, infatti, è il frutto di un processo valutativo, di giudizi di valore, che perciò non può che essere estraneo a una mente meccanica. Tale giudizio valoriale, tuttavia, diviene il prodotto di un pensiero artificiale nel momento in cui le conoscenze fornite alla macchina sono ideologicamente orientate *a priori*, in quanto tutte accomunate da un giudizio valoriale che, se non può essere elaborato dalla macchina, nella macchina è inserito.

È il problema dei pregiudizi algoritmici, definiti anche pregiudizi digitali derivati, per distinguerli dai pregiudizi digitali autonomi<sup>58</sup>. Come già esposto, il problema è emerso, tra gli altri, per l'algoritmo COMPAS, utilizzato per l'elaborazione di giudizi predittivi dalle Corti statunitensi<sup>59</sup>.

L'obiettivizzazione dei risultati dipende non tanto dal ragionamento interno alla macchina, quanto dalle informazioni in essa introdotte. Per questo la Commissione correttamente individua l'inserimento dei dati come un momento di centrale rilievo, dal quale elevati rischi possono derivare ai diritti fondamentali. In prospettiva futura, vengono individuate come prescrizioni da inserire in un testo legislativo quelle finalizzate a fornire garanzie sulla sicurezza dei prodotti, garanzie relative, pertanto, anche al carattere non discriminatorio degli esiti del pensiero artificiale dovuto a un set di dati di addestramento ristretto, che non contempli tutti gli scenari pertinenti<sup>60</sup>. All'uopo, potrebbe essere imposto come obbligatorio, a parere della Commissione, l'uso di dati sufficientemente rappresentativi.

---

<sup>57</sup> Cap. 1, par. 5 sull'algoritmo discriminatorio.

<sup>58</sup> A. Venanzoni, *La valle del perturbante: il costituzionalismo alla prova delle intelligenze artificiali e della robotica*, cit.

<sup>59</sup> Il report pubblicato dal giornale ProPublica, a firma di J. Angwin – J. Larson – S. Mattu – L. Kirchner, *Machine Bias*, in [www.propublica.org](http://www.propublica.org), 23 maggio 2016, sostiene che il COMPAS sarebbe un algoritmo discriminatorio in quanto basato su dei pregiudizi contro l'etnia afroamericana. In altri termini, esso produrrebbe dei giudizi valoriali, dunque non obiettivi.

<sup>60</sup> Libro bianco, p. 21.

Il secondo aspetto di cui il legislatore europeo dovrà occuparsi è individuato nella tenuta dei dati e dei registri<sup>61</sup>. In questo modo, si mira a ottenere le informazioni relative alla programmazione dell'algoritmo e ai dati impiegati per l'addestramento di sistemi a alto rischio. Lo scopo, come è evidente, è quello di consentire una verifica a posteriori sullo sviluppo decisionale seguito dall'algoritmo. Secondo la Commissione, i registri dovrebbero essere conservati per un periodo di tempo limitato e ragionevole e dovranno essere messi a disposizione delle autorità competenti in caso di ispezioni o ricerca di prove. Naturalmente, tale pubblicizzazione del processo produttivo si scontra con la segretezza della proprietà industriale e del diritto d'autore. Per questo, sarà necessario contemperare le due esigenze, al fine di evitare che l'interesse privato sotteso al segreto commerciale prevalga sugli interessi a una informazione completa ed esaustiva e, più in generale, alla trasparenza.

Va osservato in senso critico che un'eccessiva "burocratizzazione" del processo produttivo potrebbe non condurre ai risultati sperati. Bisognerà tenere a mente e imparare dall'esperienza maturata nel settore sanitario, ove la necessità di informare i pazienti sui risvolti di eventuali interventi e di documentare ogni scelta del medico ha portato allo sviluppo della cosiddetta medicina difensiva, con la conseguenza che l'attenzione si è spostata dal paziente ai documenti da sottoporre al paziente. Ebbene, questo rischio andrebbe evitato in nuce nel nuovo settore delle tecnologie di IA, per non dover poi trovare un rimedio successivo alle scelte di un legislatore burocrate.

Un terzo ambito di intervento, connesso a quello precedente in quanto volto a soddisfare le esigenze di trasparenza dei sistemi di IA, è rappresentato dalla previsione di obblighi di informazione<sup>62</sup>, rispondenti all'esigenza di garantire la trasparenza nella produzione e nell'addestramento degli algoritmi. Bisognerà indicare, tra le altre cose, l'obiettivo per il quale il sistema di IA è stato concepito, le condizioni in presenza delle quali esso funzionerà, nonché la precisione con la quale l'obiettivo prefissato potrà essere raggiunto<sup>63</sup>. Quest'ultimo aspetto si deve intendere riferito al *margin di errore* che l'algoritmo ha mostrato in sede di addestramento. In questo modo, l'utilizzatore, sia esso

---

<sup>61</sup> Par. D, lett. b), p. 21.

<sup>62</sup> Par. D, lett. c), p. 22.

<sup>63</sup> Libro bianco, p. 22.

un privato o un soggetto pubblico, potrà valutare l'affidabilità della risposta data dal sistema e decidere se impiegarlo in un dato contesto e quale tipo di controllo esercitare nel corso del funzionamento, dato il margine di errore reso noto dal produttore.

Viene poi in rilievo l'aspetto della robustezza e della precisione<sup>64</sup>, che si ricollega all'idea di una IA affidabile. Si richiede, difatti, che i sistemi siano sviluppati "in modo responsabile e con la debita e adeguata valutazione ex ante dei rischi che possono generare"<sup>65</sup>. La robustezza di un sistema richiede, inoltre, che i sistemi "si comportino in modo affidabile secondo le previsioni". Naturalmente si tratta di un requisito che va adattato ai sistemi di apprendimento automatico, il cui comportamento, come si è già tentato di spiegare nel capitolo precedente, non è sempre prevedibile, poiché l'algoritmo si modifica alla luce delle conoscenze acquisite nel corso del suo utilizzo. Il destinatario della normativa sarà tenuto ad adottare tutte le misure necessarie per neutralizzare o quantomeno ridurre il rischio di danni. Ciò comporta, secondo la Commissione, lo sviluppo di regole che impongano la riproducibilità dei risultati cui giunge il sistema e la gestione di eventuali errori in tutto il ciclo di utilizzo del sistema. Ma la robustezza attiene anche alla capacità del sistema di resistere ad attacchi provenienti dall'esterno, volti a manipolare i dati o l'algoritmo, come gli attacchi hacker.

Infine, l'ultimo aspetto che in questa sede preme evidenziare, è la conferma di un'impostazione antropocentrica dello sviluppo di sistemi di IA<sup>66</sup>. Il costante controllo umano è requisito ritenuto necessario dalla politica dell'Unione europea in materia di IA, come già emerso in sede di esame della Risoluzione del Parlamento Europeo. A questa idea si rifà anche la Commissione, allorché richiede che la futura legislazione imponga agli sviluppatori di sistemi di IA ad alto rischio un coinvolgimento degli esseri umani al fine di impedire che il sistema possa acquisire totale autonomia. Si suggeriscono quattro forme di controllo: *convalida*, *riesame*, *monitoraggio*, *imposizione di vincoli operativi in fase di progettazione*. Le prime tre forme di controllo avvengono al momento dell'*utilizzo* del sistema di IA. La quarta, invece, avviene nella *fase di progettazione* del sistema. Con

---

<sup>64</sup> Par. D, lett. d), p. 23.

<sup>65</sup> Libro bianco, p. 23.

<sup>66</sup> *Idem*, par. D, lett. e), p. 23.

riguardo proprio a quest'ultima modalità di controllo, l'obiettivo è quello di garantire la predeterminazione delle scelte del sistema in presenza di date condizioni<sup>67</sup>.

Passando all'analisi del controllo in fase di utilizzo del sistema, la *convalida* presuppone che il risultato del sistema di IA divenga effettivo solo dopo essere stato rivisto e convalidato da un essere umano; in questo modo la decisione definitiva è imputabile all'agente umano e non all'agente macchina.

Il *riesame* attiene a una decisione assunta dall'applicazione di IA. Tale decisione diviene immediatamente effettiva e l'intervento dell'essere umano è successivo o *ex post* e eventuale, poiché si tratta di una revisione della decisione. Ad esempio, una carta di credito viene bloccata automaticamente dal sistema, ma deve poi essere possibile il riesame di tale scelta algoritmica.

Il *monitoraggio* richiede il controllo umano durante il funzionamento del sistema, il quale ultimo può essere disattivato nel momento in cui sia riscontrato un errore. È questo il caso delle auto a guida autonoma, dove deve essere prevista una procedura di arresto attivabile dal conducente.

Si tratta di modi diversi di supervisione umana, presi in considerazione anche dal documento "Orientamenti etici per una IA affidabile"<sup>68</sup> pubblicato dal Gruppo indipendente di Esperti ad Alto Livello sulla IA. In particolare si distinguono le tre modalità di sorveglianza: *human in the loop* o approccio con *intervento* umano, *human on the loop* o approccio con supervisione umana e *human in command* o approccio con *controllo* umano. Quest'ultimo è sicuramente accostabile all'attività di *monitoraggio* di cui si dà atto nel Libro bianco. Mentre l'intervento umano (*human in the loop*) consente l'inserimento dell'uomo in ogni ciclo decisionale del sistema, quindi in fase di utilizzazione e di progettazione, per contro, l'approccio *human on the loop* prevede l'intervento umano durante il ciclo di progettazione del sistema e il monitoraggio del funzionamento.

### 1.3 La definizione di IA

---

<sup>67</sup> Per spiegare questo punto, è riportato l'esempio delle auto a guida autonoma, per le quali già nella fase di progettazione deve essere previsto che essa mantenga sempre una certa distanza di sicurezza rispetto al veicolo che lo precede.

<sup>68</sup> Precisamente a pag. 18 nella versione italiana.

A questo punto si pone come necessaria una definizione di IA, tanto per le finalità previste dal Libro bianco, che per ogni altra “futura iniziativa legislativa”<sup>69</sup>.

Una prima definizione era contenuta nella Comunicazione della Commissione sulla IA per l’Europa, COM (2018) 237<sup>70</sup>, dove si legge: *“Intelligenza artificiale” (IA) indica sistemi che mostrano un comportamento intelligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi. I sistemi basati sull’IA possono consistere solo in software che agiscono nel mondo virtuale (per esempio assistenti vocali, software per l’analisi delle immagini, motori di ricerca, sistemi di riconoscimento vocale e facciale); oppure incorporare l’IA in dispositivi hardware (per esempio in robot avanzati, auto a guida autonoma, droni o applicazioni dell’Internet delle cose)”*<sup>71</sup>.

La definizione è stata ripresa dal Gruppo di esperti ad Alto livello sull’IA, cioè lo stesso gruppo indipendente che ha elaborato le linee guida etiche di cui sopra si è riferito.

Il Gruppo di esperti ha elaborato un documento<sup>72</sup> con lo scopo di fornire una definizione di IA cui si riferiranno i documenti facenti capo a tale gruppo. La definizione aggiornata proposta dal Gruppo istituito dalla Commissione europea è la seguente: *“I sistemi di intelligenza artificiale (IA) sono sistemi software (ed eventualmente hardware) progettati dall'uomo che, dato un obiettivo complesso, agiscono nella dimensione fisica o digitale percependo il proprio ambiente attraverso l'acquisizione di dati, interpretando i dati strutturati o non strutturati raccolti, ragionando sulle conoscenze, o elaborando le informazioni derivate da questi dati e decidendo le migliori azioni da intraprendere per raggiungere l'obiettivo dato. I sistemi di IA possono usare regole simboliche o apprendere un modello numerico, e possono anche adattare il loro comportamento analizzando come l'ambiente è influenzato dalle loro azioni precedenti.*

*Come disciplina scientifica, l’IA include diversi approcci e diverse tecniche, come l’apprendimento automatico (di cui l’apprendimento profondo e l’apprendimento per rinforzo sono esempi specifici), il ragionamento meccanico (che include la pianificazione, la programmazione, la rappresentazione delle conoscenze e il ragionamento, la ricerca*

---

<sup>69</sup> White Paper, p. 16.

<sup>70</sup> Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e Sociale Europeo e al Comitato delle Regioni, Bruxelles, 25.4.2018, COM(2018) 237 final.

<sup>71</sup> *Id.*, p. 1

<sup>72</sup> Si tratta del documento dal titolo *Una definizione di IA: principali capacità e discipline*, reso pubblico il 10 aprile 2019. In esso si precisa che la definizione di IA è elaborata ai fini dei documenti riconducibili al Gruppo indipendente di esperti ad alto livello sull’intelligenza artificiale.

*e l'ottimizzazione), e la robotica (che include il controllo, la percezione, i sensori e gli attuatori e l'integrazione di tutte le altre tecniche nei sistemi ciberfisici) ”<sup>73</sup>.*

La definizione proposta dal Gruppo di esperti potremmo dire essere complessa, nel senso che essa prima individua il significato di “sistema di IA”, per poi passare alla nozione di IA come disciplina scientifica.

Seguendo una finalità esplicativa, il documento fornisce una descrizione molto essenziale e astratta di un sistema di IA, di cui di seguito si darà conto, in quanto ritenuta utile ai nostri fini.

Tre sono le capacità principali di un sistema di IA: la percezione, il ragionamento o processo decisionale e, infine, l’attuazione. Tali caratteristiche rendono il sistema di IA razionale, cioè in grado di scegliere la migliore soluzione tra quelle possibili al fine di raggiungere un dato obiettivo.

Tale processo decisionale razionale si attiva per mezzo di sensori, che dotano il sistema di IA di una capacità percettiva dei dati dall’ambiente esterno. Tali dati, una volta acquisiti, vengono trasformati in informazioni comprensibili al modulo di ragionamento e elaborazione di cui il sistema è dotato. In questo modo, il sistema perviene a una decisione eseguita mediante gli attuatori di cui dispone.

L’esempio riportato nel documento è quello del sistema di IA che pulisce il pavimento di una stanza quando è sporco. I sensori potrebbero consistere in delle fotocamere che scattino foto nel pavimento; l’immagine è il dato che viene trasmesso al modulo di ragionamento e elaborazione. Tale dato viene anzitutto trasformato in informazione attraverso un’attività di interpretazione che consente al sistema di ricondurre l’informazione entro il modello “pavimento sporco” o “pavimento pulito”. In questa prima fase si assiste a una modellizzazione o rappresentazione delle conoscenze. Segue un’attività di ragionamento basato su tali conoscenze, che porta alla scelta della condotta migliore per raggiungere l’obiettivo della pulizia del pavimento. Infine, la decisione è attuata emettendo un segnale che attiva un aspirapolvere nel caso in cui

---

<sup>73</sup> Per una diversa definizione si veda G. Ubertis, il quale riporta la definizione contenuta nella «Carta etica europea sull’utilizzo dell’intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi», adottata nei giorni 3-4 dicembre 2018 dalla Commissione europea per l’efficienza della giustizia (Cepej), G. Ubertis, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Riv. trim. dir. pen. cont.*, 4/2020, p. 76.

l'azione scelta sia quella di pulire il pavimento. Così, l'azione eseguita può modificare l'ambiente; nell'esempio, tale modifica avviene attraverso la pulizia del pavimento.

Il documento spiega anche l'apprendimento automatico, prendendo le mosse dalla distinzione tra sistemi di IA razionali e sistemi razionali che apprendono. Nel primo caso, il sistema modifica l'ambiente, ma non adatta il proprio comportamento per conseguire l'obiettivo designato in modo più efficace. Per converso, il sistema dotato di capacità di apprendimento, misura il successo della propria azione e adatta le proprie regole di ragionamento e i propri metodi decisionali, in un'ottica di efficienza.

Dopo aver messo in evidenza che la capacità di ragionamento e di apprendimento sono il risultato di *tecniche* messe in atto per la costruzione di sistemi di IA, il documento si sofferma sul tema della 'IA come disciplina scientifica'.

Con tale espressione, il Gruppo di esperti allude alle tecniche attualmente usate per costruire sistemi di IA, che vengono suddivise in due gruppi: quelle attinenti alla capacità di ragionamento e quelle che afferiscono all'apprendimento.

Le tecniche riferite alla capacità di ragionamento consentono alla macchina di svolgere tutto il procedimento sopra esposto, che dall'acquisizione dei dati, trasformati poi in conoscenze sussunte entro un certo modello, giunge alla decisione. Ebbene, tra tali tecniche la definizione di cui ci stiamo qui occupando include le attività di pianificazione, programmazione, rappresentazione delle conoscenze e ragionamento, la ricerca e l'ottimizzazione. Diverso è l'apprendimento, che, come già detto, consente alla macchina di imparare a risolvere problemi non definibili *ex ante* in modo preciso, ovvero problemi la cui soluzione non può essere sintetizzata a monte con regole di ragionamento simboliche<sup>74</sup>. In tal modo, la macchina sviluppa conoscenze autonome e ulteriori, che la condurranno a assumere decisioni imprevedibili *ex ante*.

La definizione proposta dal Gruppo di esperti è ripresa dal Libro bianco ed è ripresa nella sostanza dalla proposta di regolamento dell'aprile 2021 della Commissione. Il Libro bianco precisa che la definizione di IA deve essere al contempo *flessibile*, al fine di adattarsi al progresso tecnologico, e *sufficientemente precisa*, per poter rispondere all'esigenza di certezza del diritto<sup>75</sup>.

---

<sup>74</sup> Pag. 4 del documento del Gruppo di esperti.

<sup>75</sup> Così si legge nel White paper, a pag. 16.

La flessibilità cui la Commissione fa riferimento sembra ricordare il concetto di neutralità tecnologica<sup>76</sup>, spesso invocato laddove diritto e tecnica si incontrino<sup>77</sup>. La neutralità tecnologica si riferisce al diritto positivo e comporta che il legislatore attraverso la normazione non deve favorire determinate tecnologie o determinati prodotti rispetto ad altri, poiché diversamente interferirebbe nello sviluppo della tecnologia, condizionandola<sup>78</sup>. In altri termini, il legislatore deve osservare la tecnica e non determinarla. Una volta assunte le vesti dell'osservatore, l'approccio del diritto sarà funzionale, perché non si concentra sul "cosa", ma sul "come": non si norma l'oggetto, ma la funzione<sup>79</sup>. Nel rapporto tra la regola e il regolato, il diritto è chiamato a orientare la tecnica.

Questo è, a mio avviso, l'approccio seguito dal legislatore europeo: non condiziona la tecnica, ma la orienta, affermando la volontà di raggiungere, attraverso le norme, lo scopo di una IA antropocentrica e affidabile. In questo modo, si vuole evitare ciò che Severino definiva la tendenza della tecnica verso l'onnipotenza, in ragione del fatto che la tecnica non è neutra in quanto opera in un mondo in sé non privo di scopi: il suo scopo è il non privarsi della capacità di realizzare qualsiasi scopo<sup>80</sup>.

Ebbene, questa tendenza all'onnipotenza può essere degenerativa se si rivolta contro l'uomo stesso, che applica la tecnica e che la tecnica studia attraverso la tecnologia. È per evitare un simile epilogo che si rende necessario l'intervento del diritto positivo.

#### **1.4 La proposta di regolamento del Parlamento europeo e del Consiglio dell'ottobre 2020. L'abbandono dell'idea di una personalità giuridica del robot. L'introduzione della nuova figura dell'operatore**

Un primo coordinamento tra l'intervento del Parlamento europeo e quello della Commissione emerge dalla recente Proposta di regolamento del Parlamento europeo e

---

<sup>76</sup> In questo senso anche A. Amidei, *Intelligenza artificiale e diritto-intelligenza artificiale e product liability: sviluppi del diritto dell'Unione europea*, *Giur. it.*, 2019, 7, 1715, il quale evidenzia il cambio di prospettiva maturato in seno alle istituzioni europee in favore del principio di neutralità tecnologica, a differenza di quanto è avvenuto con le prime riflessioni sul tema, dichiaratamente finalizzate a promuovere nuovi interventi legislativi, atteso che lo stesso Parlamento esortava la Commissione a elaborare una proposta di direttiva *ad hoc* sul diritto civile e robotica.

<sup>77</sup> Irti, Severino, *Le domande del giurista e le risposte del filosofo (un dialogo su diritto e tecnica)*, in *Contratto e impresa*, 2006, p. 665.

<sup>78</sup> G. Finocchiaro, *Riflessioni su diritto e tecnica*, in *Dir. dell'informazione e dell'informatica*, fasc. 4-5, 2012, p. 831 ss.

<sup>79</sup> *Idem*, p. 833.

<sup>80</sup> Irti, Severino, *op. cit.*, p. 675.

del Consiglio sulla responsabilità per il funzionamento dei sistemi di intelligenza artificiale<sup>81</sup> datata 20 ottobre 2020.

Il documento rende più concreta l'azione del legislatore europeo, che questa volta avanza proposte dai confini più certi e stabili, con poche novità rispetto a quanto è già emerso nel dibattito europeo e molte conferme di approcci prospettati nella prima Risoluzione del Parlamento stesso e negli interventi della Commissione.

Il testo mostra la maggiore consapevolezza acquisita del legislatore europeo nel settore dei sistemi di IA e di ciò se ne ha la riprova già dalla nomenclatura utilizzata: è abbandonata l'espressione "robotica" in favore della più precisa denominazione "Intelligenza Artificiale".

La principale novità è rappresentata dall'accantonamento dell'idea di istituire una personalità giuridica del robot. Il Parlamento Europeo sembra aver preso in considerazione le osservazioni critiche mosse all'idea di istituire una personalità elettronica, tanto è vero che la proposta non affronta il tema in questione, concentrandosi unicamente sulla disciplina della responsabilità dell'operatore<sup>82</sup>.

Accanto a questo cambio di prospettiva, la Risoluzione conferma l'atteggiamento duplice della disciplina europea già proposto nella Risoluzione del 2017, ritenendo che le nuove norme sulla responsabilità civile dovranno coordinarsi con le norme sulla responsabilità da prodotto. In quest'ottica, la direttiva 85/374/CEE e la futura proposta di regolamento costituiscono "i due pilastri di un quadro comune" in materia di responsabilità per i sistemi di IA<sup>83</sup>, che dovranno essere coordinati per garantire una piena armonizzazione dei regimi di responsabilità per i sistemi di IA.

La direttiva sui danni da prodotto continua a rappresentare un testo attuale e efficace ai fini risarcitori, sebbene il Parlamento non taccia la necessità di "adattarla al mondo digitale" per continuare a garantire "un livello elevato di efficace protezione dei consumatori, come pure la certezza giuridica per i consumatori e le imprese"<sup>84</sup>.

---

<sup>81</sup> Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale (2020/2014(INL)).

<sup>82</sup> E, anzi, nota il Parlamento che "*Qualsiasi cambiamento richiesto riguardante il quadro giuridico esistente dovrebbe iniziare con il chiarimento che i sistemi di IA non possiedono né una personalità giuridica né una coscienza umana e che il loro unico compito consiste nel servire l'umanità*", par. (6) del Testo della Proposta.

<sup>83</sup> Par. 6. dell'*Allegato alla Risoluzione: Raccomandazioni dettagliate per l'elaborazione di un regolamento del parlamento europeo e del consiglio sulla responsabilità per il funzionamento dei sistemi di intelligenza artificiale, lett. A Principi e obiettivi della proposta.*

<sup>84</sup> Par. 8. della sezione *Responsabilità e intelligenza artificiale* della Risoluzione.

A tal fine, un invito è rivolto alla Commissione, alla quale è chiesto di intervenire per chiarire la definizione di “prodotti” per stabilire se vi rientrino anche i servizi digitali e adeguare i concetti di “pregiudizio” “difetto” e “produttore”. In particolare, quest’ultimo concetto dovrebbe includere i nuovi soggetti coinvolti nella produzione dei sistemi di IA; ai fini processuali, invece, l’invito è quello di valutare di invertire l’onere della prova, che, secondo il criterio della *vicinitas*, sarebbe più opportuno gravasse sul produttore e non sul consumatore.

Il coordinamento tra il nuovo regolamento e la direttiva sulla responsabilità da prodotto difettoso dovrebbe realizzarsi attraverso un regime di responsabilità integrato, con conseguenze sull’individuazione dei soggetti destinatari della nuova disciplina. La proposta di regolamento, infatti, si rivolge non al *produttore*, bensì all’*operatore*. Si tratta di una nuova figura soggettiva, che comprende tanto l’operatore di *front-end*, che di *back-end*. È “operatore di front-end”: *la persona fisica o giuridica che esercita un certo grado di controllo su un rischio connesso all’operatività e al funzionamento del sistema di IA e che beneficia del suo funzionamento*; mentre è “operatore di back-end”: *la persona fisica o giuridica che, su base continuativa, definisce le caratteristiche della tecnologia e fornisce i dati e il servizio di supporto di back-end essenziale e pertanto esercita anche un elevato grado di controllo su un rischio connesso all’operatività e al funzionamento del sistema di IA*<sup>85</sup>. L’operatore di *back-end* può coincidere col produttore: in questo caso egli sarà soggetto solo alle norme della direttiva 85/374/CEE<sup>86</sup>. Nondimeno, se vi è un solo operatore e questi coincide col produttore, le disposizioni del regolamento dovrebbero prevalere su quelle della direttiva<sup>87</sup>.

Così il Parlamento regola il rapporto tra le fonti: posto che il regolamento si rivolgerà all’operatore e posto che la direttiva 85/374/CEE si applica al produttore, le due forme di responsabilità non si applicheranno mai contestualmente all’operatore che sia anche produttore; questi risponderà secondo le norme della direttiva, se a concorrere con lui vi saranno altri operatori (non produttori) ai quali invece si applicherà la normativa sulla responsabilità da IA; mentre se vi è un solo operatore che è anche il produttore, le disposizioni del regolamento prevarranno su quelle della direttiva.

---

<sup>85</sup> Rispettivamente lett. e) e lett. f) art. 3 della Proposta, rubricato “Definizioni”.

<sup>86</sup> Questo è quanto emerge dalla lett. d) dell’art. 3, che, definendo la persona dell’operatore, recita: “d) *“operatore”*: *operatore sia di front-end sia di back-end*, a condizione che la responsabilità di quest’ultimo non sia già coperta dalla direttiva 85/374/CEE”.

<sup>87</sup> Par. (10) della *Proposta di regolamento*.

Ne consegue che la responsabilità del futuro regolamento troverà applicazione al produttore solo in via residuale.

Quella di “operatore”, pertanto, è una categoria più ampia di quella di “produttore”, perché può comprendere anche l’utente o utente della tecnologia di IA, considerato responsabile in modo analogo a quanto avviene col proprietario di un’automobile<sup>88</sup>. Il produttore e l’utente sono accomunati dalla circostanza di avere un potere di controllo del rischio associato al sistema di IA, e ciò giustifica l’attribuzione della responsabilità a loro carico. Laddove l’utente in concreto non sia qualificabile come operatore, non andrà esente da responsabilità: non risponderà secondo le disposizioni del regolamento, ma resta ferma la sua responsabilità per colpa ove abbia cagionato danni a terzi<sup>89</sup>.

Passando alla struttura del regime di responsabilità, il Parlamento accoglie la distinzione, promossa dalla Commissione, tra sistemi di IA a alto rischio e sistemi di IA non a alto rischio. Per i primi la responsabilità dell’operatore è *oggettiva* (art. 4 della Proposta di regolamento), ed è esclusa solo ove il danno o il pregiudizio siano dovuti a cause di forza maggiore. Per i sistemi non ad alto rischio l’operatore risponde a titolo di *colpa*, ma sono previste delle cause di esclusione della responsabilità. In particolare, oggetto della prova liberatoria è la dimostrazione che il sistema di IA si è attivato senza che l’operatore ne fosse a conoscenza e nonostante questi avesse adottato tutte le misure per evitare siffatta attivazione; ovvero che si sia agito con la diligenza richiesta<sup>90</sup>.

Questo regime di responsabilità differenziato conferma la scelta per *l’approccio basato sul rischio*, già proposto nella prima Risoluzione del 2017. Tale approccio presuppone l’individuazione dei criteri per distinguere i sistemi di IA in sistemi ad alto

---

<sup>88</sup> Par. 10 della sezione *Responsabilità e intelligenza artificiale* della Risoluzione e par. (9) della *Proposta di Regolamento*.

<sup>89</sup> In questo senso il par. (11) della *Proposta di regolamento*: “Se un utente, in particolare la persona che utilizza il sistema di IA, è coinvolto nell’evento dannoso, dovrebbe essere chiamato a rispondere a norma del presente regolamento solo laddove si qualifichi anche come operatore. In caso contrario, l’entità del contributo al rischio da parte dell’utente, per negligenza grave o intenzionale, potrebbe comportare la responsabilità per colpa dell’utente nei confronti del ricorrente. È opportuno che i diritti dei consumatori applicabili riferiti all’utente rimangano impregiudicati”.

<sup>90</sup> Art. 8, c. 2 della *Proposta di regolamento*: 2. *L’operatore non è responsabile se riesce a dimostrare che il danno o il pregiudizio arrecato non è imputabile a sua colpa per uno dei seguenti motivi: a) il sistema di IA si è attivato senza che l’operatore ne fosse a conoscenza e sono state adottate tutte le misure ragionevoli e necessarie per evitare tale attivazione al di fuori del controllo dell’operatore, o b) è stata rispettata la dovuta diligenza con lo svolgimento delle seguenti operazioni: selezionando un sistema di IA idoneo al compito e alle competenze, mettendo debitamente in funzione il sistema di IA, monitorando le attività e mantenendo l’affidabilità operativa mediante la periodica installazione di tutti gli aggiornamenti disponibili.*

rischio e non ad alto rischio, individuazione rimessa alla Commissione, che dovrà poi essere trasposta in un *allegato al regolamento*, sottoposto a aggiornamento periodico<sup>91</sup>. Sul punto, il Parlamento fornisce già qualche indicazione laddove afferma che l'alto rischio sussiste quando la IA si caratterizzi per il funzionamento autonomo<sup>92</sup>. Il Parlamento prende in considerazione, altresì, il caso in cui un sistema di nuova produzione non sia stato ancora classificato dalla Commissione; ebbene, in tale evenienza, il regime di responsabilità oggettiva troverà applicazione qualora il sistema abbia causato “incidenti ripetuti che producono gravi danni e pregiudizi”. Si tratta di un modello derogatorio che, a nostro avviso, dovrebbe trovare applicazione in via del tutto eccezionale.

Infine, è confermato il regime di assicurazione obbligatoria per gli operatori di sistemi di IA ad alto rischio, sul modello della responsabilità per i danni da circolazione di veicoli a motore.

Il Parlamento invita la Commissione a elaborare una proposta di regolamento ai sensi dell'art. 225 TFUE.

## **1.5 La proposta di regolamento della Commissione europea**

L'invito all'elaborazione della proposta di regolamento ai sensi dell'art. 225 TFUE è stato accolto dalla Commissione europea, che ha formalizzato un regime armonizzato di regole sull'IA<sup>93</sup>. La proposta di regolamento è dell'aprile del 2021 e, sulla scia degli atti a livello europeo che la precedono, effettua una distinzione tra sistemi ad alto e basso rischio (art. 6), per dedicarsi principalmente alla disciplina dei primi.

Il testo è composto da 85 articoli, pertanto non è certo possibile in questa sede analizzare tutti i profili della disciplina, per cui ci soffermeremo su alcuni aspetti rilevanti ai fini della nostra indagine.

Il sistema di IA è preso in considerazione in quanto prodotto o componente di un prodotto e una prima nota di rilievo va fatta alla definizione di IA di cui all'art. 3, che riprende nella sostanza quella contenuta nel documento elaborato dal Gruppo

---

<sup>91</sup> Par. 14-16 della sezione *Norme in materia di responsabilità diverse per i diversi tipi di rischi* della *Risoluzione*.

<sup>92</sup> Par. 15 della medesima sezione.

<sup>93</sup> Proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione, COM(2021) 206 final.

indipendente di esperti ad alto livello sull'intelligenza artificiale. È però una definizione più "snella" poiché non contiene la descrizione delle tecniche di produzione, per le quali si rinvia all'allegato I. Sembra allora attuarsi l'idea avanzata in sede di redazione del libro bianco di una definizione sufficientemente precisa ma al contempo *flessibile*, che può essere soggetta a integrazioni, con modifiche che, come espressamente previsto all'art. 4, interesseranno eventualmente l'allegato e non il testo principale del regolamento<sup>94</sup>.

Emerge però solo ora con chiarezza che gli obiettivi perseguiti da queste tecnologie sono indicati dall'uomo, come a voler sottolineare, già nella definizione, la necessità di un controllo umano costante sull'operato del sistema. Il tema della sorveglianza umana è approfondito al successivo art. 14, dove si prevede che i sistemi ad alto rischio siano sviluppati in modo da poter essere efficacemente supervisionati da persone fisiche durante il loro utilizzo.

Quanto al profilo soggettivo, diverse sono le figure prese in considerazione. Nella specie, scompare la figura del produttore che è sostituita da quelle del fornitore (anche di piccole dimensioni), del rappresentante autorizzato, dell'importatore e del distributore. Anche qui, come nella proposta di regolamento del Parlamento, compare la figura dell'operatore, che accosta l'utente a tutti i protagonisti della fase produttiva, confermandosi come una categoria ibrida, ma non viene riproposta la distinzione tra operatore di front-end e back-end.

La distinzione tra queste categorie soggettive scompare allorché i soggetti diversi dal fornitore commercializzano il sistema ad alto rischio con il loro nome o marchio, ovvero quando intervengano sullo stesso modificandone le finalità o apportandovi modifiche sostanziali (art. 28). In tali circostanze i distributori, importatori e persino gli utenti sono considerati come dei fornitori ai fini del regolamento e quindi soggetti ai relativi obblighi di tale categoria.

Questi obblighi riflettono le caratteristiche che la proposta di regolamento richiede per i sistemi di IA messi in commercio: i fornitori devono garantire che i sistemi di IA siano conformi ai requisiti di robustezza e accuratezza (art. 15), trasparenza (art. 13) e che i set di dati di addestramento e convalida siano "pertinenti, rappresentativi, esenti da errori e completi" (art. 10). Viene posta l'attenzione anche sul tema della cybersicurezza: i sistemi di IA ad alto rischio sono prodotti altamente tecnologici, ma al contempo

---

<sup>94</sup> Cfr. par. 5.2.1 della relazione introduttiva. La definizione di IA è contenuta all'art. 3 della proposta, che rinvia all'all. I, che fa riferimento a tutte le tecniche di produzione.

vulnerabili ed esposti ai rischi di attacchi esterni di hackeraggio, per questo nel testo si ha cura di precisare che essi devono essere resilienti in caso di manipolazioni non autorizzate da parte di soggetti terzi (art. 15, par. 4).

Il controllo del sistema ad alto rischio interessa tutto il ciclo di vita dello stesso: prima che il sistema sia immesso sul mercato deve essere sottoposto a una procedura di valutazione di conformità (artt. 19 e 43), mentre dopo il sistema è sottoposto a monitoraggio (art. 61). Ciò conferisce un ruolo di controllo anche all'utente, che viene responsabilizzato, in quanto obbligato a informare il fornitore in caso di sospetto di rischi derivanti dall'uso del sistema, che quindi dovrà essere interrotto (art. 29, par. 4).

La conformità ai requisiti indicati dal regolamento è certificata da organismi deputati all'attività di valutazione della conformità dei sistemi di IA ad alto rischio (art. 44).

Per i sistemi di IA non ad alto rischio è prevista l'elaborazione di codici di condotta da parte dei fornitori; non si tratta però di un obbligo, ma di una prassi che deve essere "incoraggiata" e "agevolata" (art. 69).

All'art. 5 è poi previsto un elenco di attività vietate. Si tratta della produzione e commercio di sistemi che utilizzano tecniche "subliminali", che sfruttano le vulnerabilità di uno specifico gruppo di persone, dell'impiego di tecniche di credit scoring (ma questo divieto vale inspiegabilmente solo nei confronti delle autorità pubbliche e non anche dei privati), l'uso di sistemi di identificazione biometrica remota "in tempo reale" (art. 5).

Le regole contenute nella proposta di regolamento si applicheranno nel territorio dell'Unione indipendentemente dalla collocazione dell'azienda produttrice. L'art. 2 prevede tra i destinatari della normativa i fornitori che immettono sul mercato europeo i loro prodotti, agli utenti situati nell'Unione, ma non solo: tra i destinatari vi sono anche i fornitori e gli utenti situati in un paese terzo, allorché l'output del prodotto sia utilizzato nell'Unione.

Come è dato notare, il regolamento rappresenta lo sviluppo di un intervento regolatorio in precedenza tracciato nelle sue aree principali nel Libro bianco, dove già si prestava attenzione al tema dei dati di input, alla tenuta dei registri, agli obblighi di informazione e alla robustezza dei sistemi di IA. Nondimeno, la proposta di regolamento delinea un quadro molto complesso nella sua fase attuativa, con oneri non indifferenti per le aziende che si potrebbero tradurre in un forte ostacolo per le piccole e medie imprese. È anche vero, però, che probabilmente tecnologie classificate come ad alto rischio saranno il prodotto di imprese di grandi dimensioni, ma ciò non fa venir meno il rischio che una legislazione di tale tipo si traduca in una regolamentazione solo simbolica,

difficile o impossibile da attuare in alcuni punti. Tra questi quello relativo ai dati: è concretamente possibile disporre di un insieme di dati esenti da errori e completi?

In generale, l'impressione che si ha è che il legislatore europeo effettui a monte una selezione dei tipi di produzione ammessi e vietati e che, disciplinando all'interno della prima categoria i sistemi ad alto rischio, si illuda di poter annullare qualsivoglia forma di pericolo. In quest'ottica può essere letta la disposizione che indica le caratteristiche dei dati sui quali l'addestramento dei sistemi di IA deve essere basato: l'individuazione di dati esenti da errori restituisce l'idea di una produzione che deve essere priva di rischi. La realtà, però, può rivelarsi diversa e mettere il legislatore europeo dinanzi all'evidenza che un insieme di dati senza nessun errore e, conseguentemente, un sistema di IA a rischio zero, è impossibile da ottenere. Questo dovrebbe far riflettere il legislatore europeo sul fatto che, disciplinando la produzione dei sistemi di IA non vietati, di fatto sta espressamente autorizzando lo svolgimento di un'attività che comporta un certo margine di rischi e che le azioni autorizzate sono in sé pericolose. Insomma, il fenomeno preso di mira non è nuovo e richiama alla mente del penalista quel compromesso tra attività socialmente utili, ancorché pericolose, e tutela dei beni giuridici che la teoria dell'*erlaubtes Risiko* ha cercato di realizzare, attraverso la delimitazione dei confini di rischio permesso o della pericolosità socialmente adeguata<sup>95</sup>.

Al legislatore europeo è quindi affidato il difficile compito di disciplinare un'attività pericolosa, favorendo la produzione di sistemi altamente tecnologici nel rispetto dei principi e dei diritti fondamentali, ma questo bilanciamento non sarà di certo realizzato per mezzo di un testo che, perfetto nella forma, è di difficile applicazione in concreto. Sembra, allora, che questo testo si basi su una contraddizione, perché da un lato autorizza la produzione e il commercio di prodotti pericolosi, dall'altro pone delle regole che sono d'ostacolo alla produzione stessa.

## **1.6 Ripartizione degli obblighi tra gli operatori economici coinvolti: dovere di diligenza e posizione di garanzia**

Un aspetto di interesse nello studio della responsabilità penale da produzione e uso dei sistemi di IA è quello della ripartizione degli obblighi tra gli operatori economici

---

<sup>95</sup> Sull'evoluzione della teoria del rischio consentito, per tutti, G. Marinucci, *La colpa per inosservanza di leggi*, Milano, Giuffrè, 1965, pp. 206 ss.

coinvolti. La tematica, in tema di reato colposo, coincide con l'individuazione del titolare del dovere di diligenza; inoltre, laddove la condotta sia omissiva, viene in rilievo la posizione di garanzia. La questione, già di per sé complessa, si presenta nel contesto della filiera produttiva in cui vi sono diverse figure, alle quali va aggiunto poi anche il consumatore del manufatto tecnologico, sotto il cui controllo si troverà il sistema dopo l'acquisto. Ben si comprende, allora, come il ciclo di vita di un sistema di IA veda coinvolti diversi soggetti, che possono essere semplificati nelle persone del produttore, del distributore o importatore e dell'utilizzatore.

Già in sede di redazione del Libro bianco la Commissione si era occupata della ripartizione degli obblighi tra gli operatori economici coinvolti<sup>96</sup>. Il criterio allora individuato era quello della *vicinitas*, secondo cui ciascun obbligo deve essere posto a carico del soggetto che si trova nella posizione migliore per affrontare eventuali rischi: bisognerà distinguere i rischi derivanti dalla fase di sviluppo dai rischi emergenti durante la fase di utilizzo del manufatto intelligente. Se, difatti, gli sviluppatori sono i più qualificati per affrontare i rischi del primo tipo, non lo saranno altrettanto nel caso in cui il rischio si verifichi al momento dell'utilizzo, per il qual caso l'obbligo dovrebbe essere stabilito a carico dell'utilizzatore<sup>97</sup>. Ragionando col linguaggio del reato omissivo improprio prima richiamato, si tratterà di capire quando vi può essere il trasferimento della posizione di garanzia da un soggetto a un altro, il quale ultimo sarà titolare di un obbligo derivato di controllo.

La più recente proposta di regolamento individua diverse categorie di operatori economici: fornitore, rappresentante autorizzato, importatore e distributore. Gli obblighi più gravosi sono quelli del fornitore, che deve garantire la conformità del prodotto ai requisiti indicati nel regolamento (art. 16) e che il sistema sia stato sottoposto alla procedura di valutazione di conformità (art. 19), deve predisporre un sistema di gestione della qualità del prodotto e deve redigere la documentazione tecnica allegata al sistema (art.16). Vi è anche un dovere di monitoraggio del sistema una volta immesso sul mercato che consente l'attuazione del dovere di informare le autorità nazionali in caso di malfunzionamento del sistema (art. 22) che possa rappresentare un rischio per la salute, la sicurezza o per la tutela dei diritti fondamentali (art. 65).

---

<sup>96</sup> Par. E, p. 25.

<sup>97</sup> *Ibidem*, p. 25.

Di certo le risposte su questo aspetto saranno diverse nel campo del diritto civile e del diritto penale. Basti qui ricordare che nel primo caso, sono ammesse forme di responsabilità oggettiva, anche dalla legislazione in tema di danno da prodotto difettoso; ebbene, come notato anche in dottrina, per i sistemi di IA non sembrano esclusi regimi di responsabilità oggettiva sul modello del danno da prodotto<sup>98</sup>. Non vale lo stesso per il diritto penale, dove le forme di responsabilità oggettiva sono fortemente osteggiate dal principio di colpevolezza<sup>99</sup>.

### 1.7 L'impiego di sistemi di IA nella giustizia penale

Tra i possibili impieghi dei sistemi intelligenti, il Parlamento europeo ha avuto cura di dedicare un'apposita risoluzione al settore della giustizia penale, con particolare riguardo all'utilizzo nelle attività di polizia<sup>100</sup>. Resta quindi secondaria la tematica della responsabilità penale da produzione e uso di manufatti intelligenti, che del resto non potrebbe essere l'oggetto principale della risoluzione, attesa la competenza penale solo indiretta dell'Unione e la potestà incriminatrice degli Stati membri in materia penale<sup>101</sup>. Nondimeno, il Parlamento non omette qualsivoglia riferimento alla responsabilità penale e le poche considerazioni espresse sul punto sono in ogni caso chiare nel confermare la necessità di una responsabilità dell'uomo, e non della macchina, anche in materia penale<sup>102</sup> e di un controllo costante durante l'utilizzo del prodotto intelligente, che deve sempre poter essere disattivato dall'uomo.

Nella consapevolezza della necessità di un sistema di responsabilità giuridica e di imputabilità per eventuali danni, il Parlamento valorizza l'attività di prevenzione di tali

---

<sup>98</sup> G. Di Rosa, *Quali regole per i sistemi automatizzati "intelligenti"?*, in *Riv. dir. civ.*, sett-ott. 2021, p. 853.

<sup>99</sup> Come noto, in Italia l'illegittimità costituzionale della responsabilità oggettiva è stata definitivamente sancita con la sentenza n. 364/1988 sull'art. 5, richiedendo che "gli elementi più significativi della fattispecie" siano coperti almeno dalla colpa. Una posizione più esplicita di rifiuto delle forme di responsabilità oggettiva è quella della successiva sentenza sul furto d'uso (Corte Cost 1085/1988), secondo cui tutti gli elementi che concorrono a contrassegnare il disvalore della fattispecie devono essere investiti dal dolo o almeno dalla colpa dell'agente. sul tema del *versari in re illicita*, per tutti, D. Pulitanò, *Diritto penale*, VIII ed. Torino, Giappichelli, p. 304 ss.

<sup>100</sup> Risoluzione del Parlamento europeo del 6 ottobre 2021 sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale (2020/2016(INI)).

<sup>101</sup> G. Grasso, *Il trattato di Lisbona e le nuove competenze penali dell'Unione europea*, in *Studi in onore di Mario Romano*, IV, Napoli, Jovene, 2011, pp. 2307 ss.

<sup>102</sup> Lett. J: "considerando che è necessario un modello chiaro per attribuire la responsabilità per i potenziali effetti nocivi dei sistemi di IA nel settore del diritto penale; che le norme regolamentari in questo ambito dovrebbero sempre sostenere la responsabilità umana e che il loro primo e principale scopo deve innanzi tutto essere la prevenzione di qualunque effetto negativo".

esiti dannosi conferendo un ruolo di rilievo al principio di precauzione nell'impiego delle tecnologie di IA per il contrasto al crimine.

Vengono sempre ribaditi i principi di trasparenza, affidabilità, robustezza e spiegabilità dei sistemi di IA, che si confermano come punti fermi del diritto dell'Unione in questo settore.

## 2. Intelligenza Artificiale e diritto civile

Il fenomeno del danno da produzione e impiego di sistemi di IA è stato affrontato più ampiamente dalla dottrina civilistica rispetto a quella penalistica, probabilmente anche perché già stimolata dal legislatore europeo.

La trattazione dell'argomento va affrontata osservando preliminarmente che secondo l'art. 1173 c.c. le fonti della responsabilità possono essere il contratto, il fatto illecito e ogni altro atto o fatto idoneo a produrre obbligazioni.

Seguendo un percorso tracciato dalla dottrina civilista italiana<sup>103</sup>, va qui osservato che nel caso di danni da sistemi di IA vengono in gioco le prime due forme di responsabilità (*ex contractu* e *ex delicto*), cui si aggiunge quella che sorge nell'ambito dei rapporti tra i professionisti che operano all'interno della catena di fornitura e produzione<sup>104</sup>, che è una responsabilità contrattuale da inadempimento, disciplinata dall'art. 1218 cc, norma che tiene conto dell'impossibilità della prestazione derivante da causa non imputabile al debitore quale condizione per l'esclusione della responsabilità debitoria. In questo contesto negoziale, il sistema intelligente viene in rilievo nella sua dimensione "statica", che considera il robot quale mero oggetto di scambio, la quale non pone problemi circa l'applicabilità della normativa tradizionale in tema di responsabilità da inadempimento<sup>105</sup>.

Diversa è la questione della responsabilità aquiliana in caso di danni causati da difetti di produzione del robot. Il caso è quello del soggetto che subisce un danno ingiusto a causa di un'azione compiuta da un sistema di IA o nella quale il sistema è coinvolto apportando un contributo.

---

<sup>103</sup> A. Santosuosso, C. Boscarato, F. Caroleo, *Robot e diritto: una prima ricognizione*, in *Giur civ.*, 2012, 7-8, 20494 ss.

<sup>104</sup> U. Ruffolo, *Intelligenza artificiale, machine learning e responsabilità da algoritmo*, in *Giur. it.*, luglio 2019, p. 1693.

<sup>105</sup> A. Santosuosso, C. Boscarato, F. Caroleo, *Robot e diritto*, cit.

Ma deve rammentarsi che, in caso di danni da produzione di sistemi di IA, accanto alle ordinarie norme aquiliane, soprattutto l'art. 2050, l'art. 2051 e comunque gli artt. 2043 ss. c.c., troverà applicazione anche la disciplina dei danni da prodotto difettoso<sup>106</sup>, ritenuta uno dei “pilastri” del regime civile per l'IA, come sostenuto dal Parlamento europeo nella Risoluzione dell'ottobre 2020<sup>107</sup>. E ciò è in perfetta coerenza con quanto previsto dall'art. 127 cod. cons., a mente del quale le disposizioni a tutela del consumatore in caso di danni da prodotto “non escludono né limitano i diritti attribuiti dal danneggiato da altre leggi”, con la conseguenza che la nostra giurisprudenza riconosce come concorrente la responsabilità da prodotto difettoso e quella aquiliana ex art. 2043 c.c.<sup>108</sup>.

Vale qui la pena rammentare che, secondo la recente Risoluzione appena sopra citata, la responsabilità per i danni da prodotto non si sommerebbe alla responsabilità civile per l'intelligenza artificiale a carico dell'operatore che dovrà essere disciplinata dal futuro regolamento. Infatti, la responsabilità del produttore sarà sottoposta alle regole di cui alla direttiva 85/374/CEE, se a concorrere con lui vi saranno altri operatori (non produttori) ai quali invece si applicherà la normativa sulla responsabilità da IA; mentre se vi è un solo operatore che è anche il produttore, le disposizioni del regolamento prevarranno su quelle della direttiva<sup>109</sup>.

## 2.1 La responsabilità del produttore

La disciplina sulla tutela del consumatore danneggiato da prodotti difettosi ha visto l'intervento del legislatore europeo con la direttiva 85/374/CEE, modificata dalla direttiva 99/34/CEE, attuata in Italia con il D.P.R. n. 224 del 1988, poi confluito nel d.lgs n. 206/2005, cioè nel codice del consumo. Per “prodotto” ai sensi dell'art. 115 cod. cons. si intende ogni bene mobile, anche se incorporato in altro mobile o immobile, incluse le energie. È invece “produttore” “il fabbricante del prodotto finito o di una sua componente, il produttore della materia prima, nonché, per i prodotti agricoli del suolo e per quelli

---

<sup>106</sup> U. Ruffolo, *Intelligenza artificiale*, cit., p. 1694.

<sup>107</sup> Par. 6. dell'*Allegato alla Risoluzione: Raccomandazioni dettagliate per l'elaborazione di un regolamento del parlamento europeo e del consiglio sulla responsabilità per il funzionamento dei sistemi di intelligenza artificiale*, lett. A Principi e obiettivi della proposta.

<sup>108</sup> U. Ruffolo, *Intelligenza artificiale*, cit. p. 1694, il quale osserva che sporadiche sono le decisioni della giurisprudenza di merito che militano in senso contrario all'applicazione dell'art. 2050 c.c. e della responsabilità da prodotto difettoso.

<sup>109</sup> Come è stato spiegato *supra*, par. *La proposta di regolamento del Parlamento europeo e del Consiglio dell'ottobre 2020*.

dell'allevamento, della pesca e della caccia, rispettivamente l'agricoltore, l'allevatore, il pescatore ed il cacciatore" (art. 115 c. 2 bis).

La responsabilità da prodotto difettoso è una responsabilità oggettiva che è fatta ricadere sul produttore senza che ne sia accertata la colpa e grava su questi anche il rischio della causa ignota: se non si conosce la causa dell'incidente, è pur sempre il produttore a rispondere<sup>110</sup>, fatte salve le cause di esclusione della responsabilità previste dalla legge all'art. 118<sup>111</sup>. Si ritiene, inoltre, che il caso fortuito non esime il produttore dall'obbligo risarcitorio, poiché la legge è silente sul punto e, a rigore, i limiti alla responsabilità del danno non possono che derivare da una espressa previsione di legge in tema di fortuito<sup>112</sup>.

Ancora, se nella catena produttiva intervengono più soggetti, come nel caso di beni di importazione, opera lo schema della responsabilità per fatto altrui<sup>113</sup>: se il produttore non è individuato, la responsabilità ricade sul "fornitore" che ha distribuito il prodotto nell'esercizio di un'attività commerciale. Ma quando il produttore opera al di fuori dell'UE, la responsabilità ricade comunque sull'importatore anche se è noto il produttore (art. 116, c. 6, cod. cons.).

Questi soggetti sono responsabili se il danno deriva dalla difettosità del prodotto. La legge specifica che è difettoso il prodotto che "non offre la sicurezza che ci si può legittimamente attendere tenuto conto di tutte le circostanze" (art. 117, c. 1, cod. cons.). Le cause del difetto possono risalire al momento della ideazione o concezione del prodotto, al processo di fabbricazione o, ancora, essere determinate dalla carenza di informazioni fornite all'utente in ordine all'utilizzo del prodotto<sup>114</sup>. In ogni caso, però, vale la regola del c.d. *rischio di sviluppo*: l'art. 118, c. 1, lett. e, cod. cons. esclude la responsabilità "se lo stato delle conoscenze scientifiche e tecniche, al momento in cui il

---

<sup>110</sup> F. Gazzoni, *Manuale di diritto privato, XVI ED.*, Napoli, Edizioni Scientifiche Italiane, 2013, p. 734.

<sup>111</sup> La norma così dispone: *1. La responsabilità è esclusa:*

*a) se il produttore non ha messo il prodotto in circolazione;*

*b) se il difetto che ha cagionato il danno non esisteva quando il produttore ha messo il prodotto in circolazione;*

*c) se il produttore non ha fabbricato il prodotto per la vendita o per qualsiasi altra forma di distribuzione a titolo oneroso, né lo ha fabbricato o distribuito nell'esercizio della sua attività professionale;*

*d) se il difetto è dovuto alla conformità del prodotto a una norma giuridica imperativa o a un provvedimento vincolante;*

*e) se lo stato delle conoscenze scientifiche e tecniche, al momento in cui il produttore ha messo in circolazione il prodotto, non permetteva ancora di considerare il prodotto come difettoso;*

*f) nel caso del produttore o fornitore di una parte componente o di una materia prima, se il difetto è interamente dovuto alla concezione del prodotto in cui è stata incorporata la parte o materia prima o alla conformità di questa alle istruzioni date dal produttore che la ha utilizzata.*

<sup>112</sup> F. Gazzoni, *Manuale*, cit., p. 734.

<sup>113</sup> A. Torrente, P. Schlesinger, *Manuale di diritto privato, XIX ed.*, Milano, Giuffrè, 2009, p. 866.

<sup>114</sup> A. Torrente, P. Schlesinger, *Manuale*, cit., p. 866.

*produttore ha messo in circolazione il prodotto, non permetteva ancora di considerare il prodotto come difettoso”.*

Questo breve esame della normativa sui danni da prodotto difettoso ci consente di individuare i due elementi su cui essa si fonda: il concetto di prodotto e la nozione di difetto.

Prendendo le mosse dalla definizione di ‘prodotto’, si riscontrano delle difficoltà nell’estendere il significante oltre il significato che il legislatore stesso gli attribuisce individuando come prodotto “ogni bene mobile, anche se incorporato in altro bene mobile o immobile”.

Da qui il dubbio circa la configurabilità dell’IA come prodotto, atteso che non è sempre chiaro se l’IA possa essere considerata un *prodotto* o un *servizio*<sup>115</sup>, distinzione rilevante in quanto a quest’ultimo non sono estensibili le norme destinate al primo. E infatti, il grado di autonomia dei sistemi di IA lascerebbe spazio a dubbi sulla classificazione degli stessi entro la categoria di prodotto, la quale dovrebbe essere rivista e ampliata, includendovi anche i servizi (atteso che l’applicazione dell’IA non riguarda solo beni, ma anche servizi)<sup>116</sup>. E proprio in quest’ottica, come già ricordato, il Parlamento ha chiesto alla Commissione di intervenire anche per precisare la nozione di “prodotto”, oltreché quella di “difetto” e “produttore”<sup>117</sup>.

L’esigenza di precisare la nozione di “difetto” sorge dalla circostanza che anche con riguardo a questo secondo elemento, su cui la responsabilità da prodotto fa perno, vi sono delle criticità, dal momento che la difettosità è stabilita sulla base dell’aspettativa di sicurezza di un consumatore medio<sup>118</sup> e che i sistemi di IA più sofisticati, cioè quelli dotati di autonomia e di capacità di autoapprendimento, pongono il problema se è considerabile “difetto” la imprevedibilità della decisione.

Per questo, in attesa di un intervento di integrazione da parte del legislatore europeo, è stato sostenuto che le norme in tema di responsabilità del produttore potrebbero trovare applicazione solo ove si consideri l’IA come oggetto e tale non sarebbe il sistema dotato

---

<sup>115</sup> Report from the Expert Group on Liability and New Technologies, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 28.

<sup>116</sup> G. Capilli, *Responsabilità e robot*, in *Nuova Giur. Civ.*, 2019, 3, p. 621 ss.

<sup>117</sup> Par. 8. della sezione *Responsabilità e intelligenza artificiale* della Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l’intelligenza artificiale (2020/2014(INL)).

<sup>118</sup> Report from the Expert Group on Liability and New Technologies, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 28.

di capacità cognitive e di adattamento all'ambiente esterno, che modifichi il suo comportamento sino a poter giungere a una soluzione imprevedibile da parte del produttore, in quanto determinata dall'esperienza fatta dall'algoritmo nel corso del suo impiego<sup>119</sup>.

Nel dibattito che si è registrato sul punto, pare più attenta l'opinione di chi ritiene che per gli *smart products* dotati di autoapprendimento la difettosità si manifesterebbe nella condotta deviante del prodotto intelligente dovuta all'addestramento ricevuto o alle esperienze successive alla sua commercializzazione<sup>120</sup>. In questi casi il difetto ben potrebbe consistere nell'assenza di "blocchi a tali evoluzioni malevole", con la evidente conseguenza che il difetto, così inteso, sarebbe genetico e solo apparentemente sopravvenuto<sup>121</sup>.

Tale impostazione giunge correttamente a isolare un nuovo centro di imputazione della responsabilità nel soggetto che crea l'algoritmo di *machine learning*, distinguendolo dalla posizione di chi acquista l'algoritmo per inserirlo in un prodotto robotico, sempreché si tratti di un soggetto diverso. Le ricadute dell'individuazione di una "responsabilità da algoritmo" sarebbero, da un lato, quella di ridurre la responsabilità di chi produce il bene finito che incorpora l'algoritmo; dall'altro, però, si aggrava l'onere probatorio in capo al consumatore. Questi, infatti, già normalmente chiamato a provare tutti gli elementi del danno da prodotto difettoso, è in tale contesto penalizzato dall'ulteriore onere di individuare il soggetto, persona fisica o giuridica, che l'algoritmo ha ideato e di provare che il difetto sia riconducibile all'opera da questi prestata nella catena produttiva. Così, ciò che dovrebbe agevolare la finalità riparatoria della responsabilità civile, cioè l'individuazione di un ulteriore soggetto chiamato a rispondere del danno, si trasmuta in un ulteriore onere probatorio in capo al consumatore danneggiato, con le conseguenze anche in punto di costi delle spese legali. La questione dell'aggravamento dell'onere probatorio del consumatore, che potrebbe frustrare le esigenze riparatorie della responsabilità da prodotto, è nota alle istituzioni europee, tanto è vero che il Parlamento, nella Risoluzione dell'ottobre 2020, ha invitato la Commissione a prevedere una inversione dell'onere probatorio per i danni causati dalle tecnologie

---

<sup>119</sup> A. Santosuosso, C. Boscarato, F. Caroleo, *Robot e diritto*, cit.

<sup>120</sup> U. Ruffolo, *Intelligenza artificiale*, cit. p. 1697, nonché A. Amidei, *Intelligenza artificiale e diritto – Intelligenza artificiale e product liability: sviluppi del diritto dell'Unione europea*, in *Giur. it.*, 2019, 7, p. 1715 ss.

<sup>121</sup> U. Ruffolo, *Intelligenza artificiale*, cit. p. 1697.

digitali emergenti, seppur “in casi previamente definiti e previa un’adeguata valutazione”<sup>122</sup>. In questo modo, l’auspicata modifica alla direttiva sulla responsabilità del produttore risolverebbe il problema della disincentivazione del ricorso alla tutela riparatoria da parte del consumatore che subisce un danno.

Ma allo stato attuale, l’onere della prova continua a gravare sul consumatore. Si è osservato che il problema dell’eccessivo aggravio in termini probatori del consumatore sarebbe temperato dal carattere cumulativo e non alternativo della responsabilità dei soggetti che intervengono nella catena produttiva<sup>123</sup>. Pertanto, la responsabilità da algoritmo si sommerebbe a quella da produzione del sistema di IA nel suo complesso inteso, secondo lo schema della responsabilità solidale, per cui il consumatore potrebbe proporre azione contro chiunque abbia apportato un contributo nella realizzazione del prodotto compravenduto, fatta salva l’azione di regresso nei rapporti interni<sup>124</sup>.

Naturalmente va qui precisato che siffatta soluzione, propria del sistema civilistico delle obbligazioni solidali, non è trasponibile sul piano penale, dove il principio di personalità della responsabilità e la tipologia della sanzione impongono che sia individuato il soggetto che ha contribuito con la propria condotta alla realizzazione dell’evento. In un caso del genere, la catena produttiva costituisce un momento di cooperazione di più soggetti che potrebbe dar luogo a una responsabilità a titolo di cooperazione colposa secondo il meccanismo dell’art. 113 c.p., in ragione del nesso subiettivo che connette le varie condotte di coloro che intervengono nella catena produttiva<sup>125</sup>. Ma ciò comporta che tutti i soggetti saranno responsabili a seconda del contributo conferito alla realizzazione dell’evento, e non certo uno solo per tutti; inoltre la responsabilità dovrà essere provata secondo i criteri di accertamento della colpa e non desunta in modo automatico.

Tornando alla responsabilità civile, va notato che, anche a individuare il soggetto responsabile e il nesso eziologico tra danno e evento, il produttore potrebbe giovare della causa di esonero da responsabilità del rischio da sviluppo. Ma è anche vero che il rischio da sviluppo, mentre esclude la responsabilità da prodotto difettoso, potrebbe rendere

---

<sup>122</sup> Par. 8. della sezione *Responsabilità e intelligenza artificiale* della Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l’intelligenza artificiale (2020/2014(INL)).

<sup>123</sup> A. Amidei, *Intelligenza artificiale e diritto*, cit.

<sup>124</sup> A. Amidei, *Intelligenza artificiale e diritto*, cit.

<sup>125</sup> È questo nesso psicologico tra le condotte a distinguere la cooperazione colposa dal concorso di cause indipendenti. Cfr. Fiandaca, Musco, *Diritto penale, parte generale*, VI ed., Bologna, Zanichelli, p. 575.

applicabile la disciplina della responsabilità da attività pericolosa, considerato che l'assenza di sufficienti conoscenze tecniche al momento della commercializzazione, conoscenze poi sopravvenute quando il prodotto è ancora sul mercato, potrebbe costituire "indice, se non addirittura prova, della pericolosità dell'attività connotata da tale rischio"<sup>126</sup>. In questo modo il consumatore danneggiato non sarebbe privato della tutela risarcitoria, che gli sarebbe per contro accordata a mente dell'art. 2050 c.c.

È anche vero che l'art. 2050 c.c. sull'esercizio di attività pericolosa si ritiene invocabile non solo nel caso in cui la responsabilità da prodotto difettoso sia esclusa per la sussistenza di una delle cause di cui all'art. 118 cod. cons., ma anche in aggiunta alla responsabilità da prodotto<sup>127</sup>. Ma i due tipi di responsabilità non saranno sempre cumulabili: il raggio d'azione dell'art. 2050 non è limitato ai soli prodotti difettosi, poiché si estende anche ai beni dotati di potenzialità lesiva, ancorché privi di difetti, cosicché ha un ambito di applicazione più esteso.

## 2.2 La responsabilità dell'utente

V'è da domandarsi se, oltre alla responsabilità per esercizio di attività pericolosa, possano operare altri schemi di responsabilità previsti nel nostro codice civile quali paradigmi normativi di allocazione del rischio laddove il danno non derivi da un difetto del robot, ma da un suo comportamento non impostato come standard dal produttore<sup>128</sup>. E infatti, la responsabilità del produttore dovrebbe essere limitata ai danni provocati da difetti di produzione o da comportamenti impostati in origine come standard: le capacità acquisite attraverso l'utilizzo devono essere tenute distinte da quelle attribuibili al produttore<sup>129</sup>. Per questo l'utilizzatore è considerato un soggetto che va responsabilizzato e ci si è interrogati sulle norme che potrebbero fondarne la responsabilità.

A livello europeo, la più recente proposta del Parlamento considera anche l'utilizzatore come soggetto destinatario di obblighi di diligenza, in quanto si tratta di una figura che ben può rientrare nella più ampia categoria di "operatore"<sup>130</sup>. Si precisa che, laddove non vi sia coincidenza tra la figura dell'operatore, così come definito dal

---

<sup>126</sup> U. Ruffolo, *Intelligenza artificiale*, cit. p. 1697.

<sup>127</sup> U. Ruffolo, *Intelligenza artificiale*, cit. p. 1694.

<sup>128</sup> A. Santosuosso, C. Boscarato, F. Caroleo, *Robot e diritto*, cit.

<sup>129</sup> A. Santosuosso, C. Boscarato, F. Caroleo, *Robot e diritto*, cit.

<sup>130</sup> Par. (11) della *Proposta di regolamento*.

legislatore, e l'utente, questi sarà comunque responsabile se ha contribuito al rischio che ha cagionato il danno a terzi<sup>131</sup>, secondo le norme sulla responsabilità civile previste dal diritto interno.

Nel nostro ordinamento potrebbe invocarsi la responsabilità del custode di cui all'art. 2051 c.c. Invero, attorno a tale soluzione si registrano opinioni divergenti. In senso affermativo, si evidenzia che la norma potrebbe fondare la disciplina della responsabilità non del produttore genericamente inteso, bensì dell'addestratore dell'algoritmo di *machine learning*. Si richiama, al riguardo, un argomento storico, che fa leva principalmente sulle origini della norma in esame nell'ordinamento francese dove, accanto alla *culpa in custodiendo*, l'art. 1384 del *Code Napoléon* è diventato base normativa per la responsabilità "per fatto della cosa", venendo a regolare anche i danni da cose dotate di una propria pericolosità (come le automobili) e da cose rese pericolose dall'uso che delle stesse fosse fatto, conoscendo così un'applicazione particolarmente estesa<sup>132</sup>.

Altri, invece, escludono il richiamo della responsabilità da cose in custodia, evidenziando che la "cosa" di cui all'art. 2051 c.c. è una cosa inanimata, per cui non vi rientrerebbe l'applicazione munita di motilità e operatività; ancora, un eventuale compito inappropriato demandato al sistema sarebbe comunque causalmente riconducibile al fattore umano che lo ha determinato, senza invocare un rapporto con la cosa, pertanto varrebbe il richiamo alla clausola generale di cui all'art. 2043 c.c. E, ancora, la responsabilità del custode non sarà invocabile allorché la cosa stessa sia stata la causa efficiente e non una sua improvvida manutenzione o un inappropriato impiego<sup>133</sup>.

Ma è anche vero che a lungo si è dibattuto in dottrina sulla possibilità di estendere la responsabilità delle cose in custodia anche ai danni cagionati da cose caratterizzate da un intrinseco dinamismo o limitarla alle sole cose inerti, per giungere a ritenere che rileva che il danno sia esplicitazione della concreta idoneità al nocimento della cosa e che quindi derivi dalla cosa stessa<sup>134</sup>. Ebbene, questa idoneità ben potrebbe essere propria di sistemi intelligenti con capacità di autoapprendimento.

---

<sup>131</sup> *Idem*.

<sup>132</sup> U. Ruffolo, *Intelligenza artificiale*, cit. pp. 1699-1700.

<sup>133</sup> M. Costanza, *L'intelligenza artificiale e gli stilemi della responsabilità civile*, in *Giur. It.*, luglio 2019, p. 1687.

<sup>134</sup> A. Thiene, art. 2051 cc. in G. Cian, A. Trabucchi, *Commentario breve al codice civile, XIV ed.*, CEDAM, Wolters Kluwer, 2020, p. 2231.

Altra norma di riferimento potrebbe essere la responsabilità del proprietario di animali *ex art. 2052 c.c. o, ancora l'art. 2048 c.c.* laddove il robot o l'algoritmo sia assimilato al minore che apprende durante il suo percorso di crescita<sup>135</sup>, con il problema, in quest'ultimo caso, di un gap di responsabilità per il caso in cui il danno derivi da una scelta compiuta dal sistema in quanto appresa dall'utilizzatore, ma posta in essere in un momento in cui l'applicazione di IA non era sotto la vigilanza dell'utilizzatore stesso<sup>136</sup>. Se il sistema non è sottoposto alla vigilanza di un terzo, cui è stato delegato il compito di vigilare sulla cosa, l'utilizzatore potrà rispondere, ma è sempre fatto salvo il caso fortuito. Tuttavia l'opinione è allo stato minoritaria<sup>137</sup>.

Analizzando il momento della responsabilità, tanto civile quanto penale, emerge in ogni caso la necessità che la relativa disciplina (avente finalità ripristinatoria, in sede civile, e sanzionatorio-rieducativa, in ambito penale) sia preceduta da una tutela preventiva che trovi spazio in sede di progettazione e sviluppo, al fine di diminuire i potenziali danni arrecati a persone o beni.

### **3. Responsabilità aquiliana e responsabilità penale: fatto illecito e fatto costituente reato**

L'analisi della normativa di diritto civile non è superflua in un contesto in cui oggetto dell'indagine è la responsabilità penale da produzione e uso di sistemi intelligenti se si considera che i diversi settori dell'ordinamento comunicano tra di loro attraverso il caso concreto, il quale può assumere la veste di fatto illecito dannoso e di reato. È pur vero che queste due branche dell'ordinamento sono poste a presidio di interessi diversi e rispondono a scopi diversi, ma va considerato che anche le regole dell'illecito civile rappresentano una forma di tutela della vittima e per questo, prima di trattare di tutela in ambito penale, era doveroso discutere anche se brevemente di quella civile.

Va altresì considerato che le differenti finalità della tutela civile la rendono più elastica rispetto a quella penale: lo scopo è riparatorio e non sanzionatorio<sup>138</sup>. Così, mentre

---

<sup>135</sup> A. Santosuosso, C. Boscarato, F. Caroleo, *Robot e diritto*, cit.

<sup>136</sup> A. Santosuosso, C. Boscarato, F. Caroleo, *Robot e diritto*, cit.

<sup>137</sup> Nega che la norma sulla responsabilità dei genitori, dei tutori, dei precettori e maestri d'arte sia estensibile alle intelligenze artificiali U. Ruffolo, *Intelligenza artificiale*, cit., il quale esclude altresì il ricorso all'art. 2046 sulla responsabilità dei padroni e committenti e dell'art. 2052 cc sul danno cagionato da animali, in quest'ultimo caso seguito anche dall'opinione di G. Capilli, *Responsabilità e robot*, cit.

<sup>138</sup> Fanno eccezione i casi di danni punitivi, ammessi nel nostro ordinamento nei limiti in cui ciò sia disciplinato dal legislatore, come precisato a partire dalla sent. Sez. Un. n. 16601 del 5 luglio 2017.

la responsabilità civile ha una prospettiva vittimocentrica, quella penale persegue uno scopo rieducativo e generalpreventivo. Questo spiega anche la storia dell'evoluzione della responsabilità aquiliana che ha visto l'ampliamento degli spazi di tutela della vittima sul piano della disciplina sostanziale e processuale.

Sul piano sostanziale si è assistito all'evoluzione della nozione di danno ingiusto che non è più inteso come lesione del solo diritto soggettivo assoluto, ma anche di diritti relativi, cioè i diritti di credito, e ingloba oggi persino posizioni giuridiche diverse dai diritti, cioè interessi legittimi, aspettative e anche la *chance*. Questa flessibilizzazione dell'oggetto di tutela si riflette sull'accertamento dei presupposti della responsabilità, in particolare del nesso di causalità.

Come noto, anche il diritto civile si avvale delle norme di cui agli artt. 40 e 41 c.p., ma esse vengono lette alla luce della teoria della causalità adeguata, che impiega regole meno stringenti rispetto al modello delle leggi scientifiche della responsabilità penale. Sul piano processuale, diverso è poi anche il parametro per ritenere accertato tale nesso, che è quello del *più probabile che non*, criterio molto più elastico rispetto a quello penalistico dell'*oltre ogni ragionevole dubbio*. Inoltre, mentre nel diritto civile l'onere della prova può essere distribuito in base al criterio della *vicinitas*, nulla di simile è prospettabile nel processo penale, dove è l'accusa a dover dimostrare la responsabilità dell'imputato<sup>139</sup>.

Ma abbandoniamo il terreno processuale per ritornare a quello del diritto sostanziale: l'evoluzione registratasi nella dottrina e nella giurisprudenza civilistiche hanno condotto all'ampliamento del risarcimento del danno non patrimoniale, oggi accordato oltre le ipotesi di responsabilità da reato, in quanto riconosciuto ogni qualvolta si sia in presenza di un interesse costituzionalmente tutelato. Infine, la già ricordata tendenza ad abbandonare forme di responsabilità basate sulla colpa a vantaggio di forme di responsabilità oggettiva segna ulteriormente un distacco tra il rimedio civile e quello penale.

Tutto ciò comporta che, collocandoci nell'ottica della responsabilità del produttore per i danni da prodotto intelligente difettoso, laddove la vittima riceverà una riparazione in termini economici per i danni subiti, non sempre il responsabile sarà destinatario di una sanzione penale. Come si avrà modo di approfondire più avanti, il diritto penale incontra

---

<sup>139</sup> Per un confronto tra gli orientamenti della giurisprudenza civile e di quella penale sui temi della causalità, ma anche della colpa, si rinvia a R. Blaiotta, *Causalità e colpa: diritto civile e diritto penale si confrontano*, in *Cass. Pen.*, 1, 2009, pp. 78 ss.

molte resistenze dovute alla sua natura frammentaria e alla clausola della *extrema ratio*; ulteriori resistenze circondano poi la categoria del diritto penale d'evento, basti pensare che l'accertamento del nesso causale avviene sulla base di parametri più rigidi, indissolubilmente legati alla regola probatoria dell'oltre ogni ragionevole dubbio. Tutto ciò non è detto che precluda a priori qualsivoglia tutela da parte del diritto penale classico di fronte al dirimente avanzare dei nuovi rischi da tecnologie intelligenti, per questo andremo a immergerci a breve in questa indagine, in cui tenteremo di rispondere al quesito "c'è ancora spazio per un diritto penale dell'evento?". Si tratta di un'indagine non nuova: tra i contributi più illustri v'è quello di Federico Stella, che si è occupato dei temi della modernità e delle tutele offerte anche da branche del nostro ordinamento diverse da quella penale<sup>140</sup>. Stella considera quella civile e amministrativa come la forma di tutela effettiva delle vittime della modernità. La riflessione va già da ora condivisa alla luce di queste brevi riflessioni sulla responsabilità da fatto illecito.

#### **4. Diritto e tecnologia: regole di convivenza**

Lo scenario inedito dei sistemi di IA apre il campo a nuovi spazi di regolamentazione, ma non necessariamente a una regolazione del tutto nuova. Nuove leggi non appaiono sempre indispensabili a fronte di un nuovo fenomeno. In questo senso è orientato il dibattito presso le istituzioni europee in punto di responsabilità per i sistemi di IA, che giunge sempre a riconoscere l'attualità della disciplina già presente, soprattutto quella in materia di danni da prodotto difettoso.

Ma si considera anche l'opportunità di introdurre nuove norme che tengano conto delle peculiarità che caratterizzano i sistemi intelligenti, in ragione della capacità di autoapprendimento.

La spinta alla regolamentazione del legislatore europeo tuttavia è sempre accompagnata dalla consapevolezza di dover porre dei limiti alla potestà regolativa del potere legislativo, nazionale e europeo, nel tentativo di non ostacolare lo sviluppo economico.

A tal proposito, in dottrina v'è chi valuta l'opportunità di affiancare alle norme di *product liability* anche discipline tecniche che possano orientare la fase produttiva, sulla

---

<sup>140</sup> F. Stella, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, Milano, Giuffrè, 2003.

scia del modello regolatore della “Direttiva Macchine” (Direttiva 2006/42/CE) o della Direttiva 2001/95/CE in materia di sicurezza dei prodotti<sup>141</sup>. In altri termini, si vuole che il legislatore, a ogni livello, sia nazionale che europeo, si preoccupi non solo del momento della responsabilità, cioè delle *issues of liability*, ma anche delle regole in materia di prevenzione dei danni stabilendo delle condizioni di produzione (*issues of permittance*) con norme a elevato contenuto tecnico che impongano standard, divieti, processi di controllo e monitoraggio, obblighi di revisione<sup>142</sup>.

L’idea di un intervento legislativo sia a monte, volto a orientare la produzione, che a valle, volto a tutelare chi patisce dei danni, è pienamente conforme al ruolo duplice che viene riconosciuto alla responsabilità, la quale deve certamente garantire un giusto ristoro a chi soffre un pregiudizio, ma al contempo deve costituire un incentivo a evitare *ab origine* di causare danni o pregiudizi<sup>143</sup>.

Opportunamente si è osservato che il tema della responsabilità nell’ambito della IA è forse il più delicato dal punto di vista giuridico, poiché esso involge anche profili legati alla crescita economica e allo sviluppo della ricerca<sup>144</sup>. E, in effetti, il settore della IA costituisce un banco di prova per la convivenza tra il diritto e la tecnologia, una convivenza in cui bisogna tracciare dei limiti non solo alla tecnologia, da parte del diritto, ma anche al diritto, da parte della tecnologia. Va quindi trovato un equilibrio tra regolazione e industrializzazione: eventuali norme nuove non devono tradursi in un ostacolo alla produzione e alla fruizione di nuove tecnologie. Nel rapporto tra tecnologia e diritto, l’intervento di quest’ultimo non deve mai apparire una penalizzazione del tutto ingiustificata nei confronti del progresso, soprattutto se si considera che proprio il progresso può essere orientato al perseguimento del miglioramento delle condizioni di vita degli esseri umani e, in taluni casi, può persino ridurre l’esposizione a pericolo di beni o interessi tutelati dall’ordinamento<sup>145</sup>.

---

<sup>141</sup> A. Amidei, *Intelligenza artificiale e diritto*, cit.

<sup>142</sup> *Idem*.

<sup>143</sup> In questo senso si esprime la Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l’intelligenza artificiale (2020/2014(INL)), par. A.

<sup>144</sup> R. Bifulco, *Intelligenza artificiale, internet e ordine spontaneo*, cit., p. 389.

<sup>145</sup> U. Ruffolo, E. Al Mureden, *Intelligenza artificiale e diritto. Autonomous vehicles e responsabilità nel nostro sistema e in quello statunitense*, in *Giur. It.*, 2019, 7, p. 1704 ss. Le osservazioni degli A. si concentrano in particolare sulle auto a guida autonoma, le quali si rivelerebbero capaci di assicurare maggiore sicurezza e diminuire significativamente i sinistri da circolazione su strada e garantirebbero altri effetti virtuosi, quali un maggiore accesso alla circolazione anche per i soggetti disabili.



### CAPITOLO III

#### INTELLIGENZA ARTIFICIALE E RESPONSABILITÀ PENALE. LA TENUTA DELLE CATEGORIE TRADIZIONALI DEL DIRITTO PENALE D'EVENTO DI FRONTE AI NUOVI RISCHI

Bisogna a questo punto passare alla trattazione dei risvolti penalistici della responsabilità da danno da sistemi di IA, per comprendere il modo in cui il complesso delle norme di diritto pubblico che prevedono fatti illeciti, per i quali sono comminate conseguenze penali, si atteggia in presenza di una condotta in cui vi sia un contributo, almeno parziale, di un sistema di IA.

Il tema presenta degli aspetti di novità ma sarebbe errato ritenerlo nuovo in ogni suo profilo. Come cercherò di dimostrare, molti interrogativi sono già stati affrontati in altre sedi e in altri contesti, tutti però connessi alla generale categoria del delitto colposo e del diritto penale del rischio, cosicché sembra prendere forma una risposta positiva al quesito posto da quella dottrina che, nell'affrontare alcuni temi di diritto penale in cui è coinvolta l'intelligenza artificiale, esordiva già dal titolo con la domanda “quesiti nuovi e soluzioni antiche?”<sup>1</sup>.

Eppure una certa dottrina non ha esitato a proporre soluzioni nuove per i nuovi quesiti sollevati dall'ingresso dell'IA nella vita quotidiana e allora bisogna individuare i possibili itinerari di responsabilità percorribili.

Degli spunti giungono dalla circostanza che da tempo il diritto penale si rivolge non più solamente alla persona fisica del reo, ma anche agli enti, quali soggetti inanimati. Questo comporta che così come la società viene considerata in alcuni ordinamenti e secondo talune opinioni destinataria del precetto penale, anche un sistema di IA potrebbe esserlo.

Per quanto concerne l'ordinamento penale italiano, va ricordato come esso abbia subito da sempre l'influenza del diritto romano, secondo cui *societas delinquere non potest*. Per conciliare la realtà empirico criminologica, che dimostra che *societas saepe delinquit*, ad esempio allorché i reati dei rappresentanti sono espressione della volontà assembleare<sup>2</sup>, il legislatore italiano ha optato per una forma di responsabilità del tutto

---

<sup>1</sup> S. Quattrococo, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale “predittiva”*, *Cass. pen.*, 4, 2019, pp. 1748 ss.

<sup>2</sup> F. Mantovani, *Diritto penale, Parte generale, X ed.*, Milano, Wolters Kluwer, CEDAM, 2019, p. 111.

peculiare, sulla cui natura giuridica, in ragione di tale peculiarità, ancor oggi si discute<sup>3</sup>. Si tratta della responsabilità degli enti da reato altrui, introdotta con il d.lgs. 231/2001.

Orbene, in ragione di questa apertura dell'ordinamento verso forme di responsabilità (penale?) delle persone giuridiche, si potrebbe sostenere l'estensione della responsabilità penale anche a altri soggetti "artificiali" e, nel caso di nostro interesse, anche ai sistemi di IA. L'opinione non è priva di sostenitori in dottrina<sup>4</sup>.

In questo modo, il letto su cui scorre questo immaginario fiume che sfocia nel mare della responsabilità per i danni da produzione e uso dei sistemi di IA si dividerebbe, potendo intraprendere più direzioni che compongono una foce a delta. Oltre al classico itinerario della responsabilità civile da fatto illecito dannoso, della responsabilità amministrativa (se legislativamente prevista), della responsabilità penale *personale* e della responsabilità da reato degli enti (collocata, a seconda dell'opinione cui si aderisce, all'interno della seconda figura di responsabilità o dell'ultima), si aggiungerebbe un nuovo itinerario, cioè quello della responsabilità *singolare* del sistema di IA, inteso quale soggetto destinatario della norma penale<sup>5</sup>. Definisco questa forma di responsabilità "singolare" perché, tanto l'aggettivo "*personale*", quanto quello "*individuale*" richiamano una soggettività umana, della persona o dell'individuo, appunto, che, a mio avviso, non si adatta allo scenario della responsabilità penale del *sistema* di IA. Quest'ultimo non è una persona, né tanto meno un individuo, ma è certamente un soggetto la cui responsabilità viene considerata, secondo questa opinione, *singularmente*, cioè *uti*

---

<sup>3</sup> Tre le opinioni sul campo. Secondo un primo indirizzo, la responsabilità ha natura amministrativa e si invoca a tal proposito l'esplicita qualificazione da parte della legge delega e del decreto legislativo, il regime della prescrizione, diverso da quello penalistico, e, ancora, la circostanza che in caso di trasformazione, fusione o scissione dell'ente il trattamento sanzionatorio richiama la disciplina civilistica sulla modificazione delle obbligazioni delle società. Un secondo indirizzo sostiene la natura penale della responsabilità in ragione della necessaria commissione di uno dei reati indicati dalla legge come presupposto della responsabilità e soprattutto, facendo leva sulla competenza del giudice penale e sull'applicazione delle regole (e delle garanzie) del diritto penale. Infine, la tesi del *tertium genus*, che fa leva sulle peculiarità di tale regime di responsabilità alla luce delle ragioni valorizzate dalle due tesi contrapposte appena citate. F. Mantovani, *Diritto penale, op. cit.*, p.114.

<sup>4</sup> La tesi fa principalmente capo a G. Hallevy, il quale ha dedicato un'ampia monografia all'argomento, dal titolo *Liability for Crimes and Artificial Intelligence*, Springer, 2014. La tesi è altresì esposta dall'A. in *The Basic Models of Criminal Liability of AI Systems and Outer Circles*, 11 giugno 2019, disponibile al sito *ssrn.com.*; *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, *Akron Intellectual Property Journal*, Vol. 4 : Iss. 2 , Article 1, 2010; *When Robots Kill: Artificial Intelligence under Criminal Law*, Northeastern University Press, 2013. Più di recente anche F. Lagioia, G. Sartor, *AI Systems Under Criminal Law. A Legal Analysis and A Regulatory Perspective*, in *Phil.&Techn.*, 2019, p. 1 ss.

<sup>5</sup> Cfr. in ambito filosofico, R. Kurzweil, *La singolarità è vicina*, Apogeo Education - Maggioli Editore, 2008.

*singulus* poiché il sistema di IA è inteso quale autonomo centro di imputazione e ciò fa di questa responsabilità un fenomeno giuridico del tutto nuovo.

Seppur in modo sintetico, è bene proporre alcune osservazioni sulle ragioni che portano ad escludere un simile itinerario di responsabilità.

La personalità dell'autore è uno dei tre pilastri su cui poggia il diritto penale moderno, assieme al fatto e alle conseguenze<sup>6</sup>. Infatti, sull'autore si concentra tutta l'indagine sul profilo psicologico della responsabilità, che consente non solo di attribuire psicologicamente il fatto all'autore, ma anche di graduare la pena, per effetto del parametro di cui all'art. 133, n. 3, c.p.. Ritenere autore di reato un sistema di IA porrebbe il problema di come accertare la *coscienza e la volontarietà* dell'azione richiesti dall'art. 42, c.1, c.p., e, ancor prima, dell'esistenza delle stesse<sup>7</sup>, col rischio di superare la dimensione soggettiva dell'illecito penale.

Ma la critica all'opinione che sostiene una responsabilità penale dei sistemi di IA mi sembra ancor più fondata se si riflette sulle conseguenze penali, cioè sulla sanzione. Essa non potrà svolgere la sua funzione preventivo-rieducativa se non quando è destinata a un soggetto che è in grado di comprendere il disvalore della propria condotta e che possa trarre beneficio dalla pena. Siffatta consapevolezza difetterebbe a un sistema di IA<sup>8</sup>, che non "capirebbe" le ragioni della sanzione e che non potrebbe per questo essere rieducato<sup>9</sup>. Ma affrontare il tema della rieducazione della pena comporta inevitabilmente un collegamento col principio di colpevolezza e quindi ritorniamo alla prima obiezione mossa alla teoria della responsabilità penale dei sistemi di IA. E infatti, come affermato dalla Consulta, la pena può svolgere la funzione rieducativa assegnatale dalla Costituzione solo se si rivolga a un soggetto *colpevole*, che abbia posto in essere quella condotta con una partecipazione soggettiva al fatto<sup>10</sup>.

---

<sup>6</sup> F. Mantovani, *Diritto penale, op. cit.*, p. XXIII

<sup>7</sup> A. Cappellini, *Machina delinquere potest? Brevi appunti sull'intelligenza artificiale e la responsabilità penale*, *Discrimen*, 27.03.2019; C. Piergallini, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, in *Riv. it. dir. proc. pen.*, 4, 2020, pp. 1745-1774, in particolare p. 1769.

<sup>8</sup> S. Gless, E. Silverman, T. Weigend, *If Robots Cause Harm, Who Is to Blame? Self-Driving Cars and Criminal Liability*, (January, 29, 2016), SSRN: <http://dx.doi.org/10.2139/ssrn.2724592>, p. 417.

<sup>9</sup> A. Cappellini, *op. cit.*; I. Salvadori, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in *Riv. it. dir. proc. pen.*, 1, 2021, pp. 83- 118, in particolare a p. 98.

<sup>10</sup> Corte Cost. sent. n. 364 del 1988. Nella sentenza i giudici sottolineano il collegamento tra il primo e il terzo comma dell'art. 27, osservando che la funzione rieducativa della pena postula almeno la colpa dell'agente in relazione agli elementi più significativi della fattispecie tipica; diversamente sarebbe contraddittorio e illogico richiedere la rieducazione di colui che, non essendo almeno "in colpa" rispetto al fatto, non ha bisogno di essere rieducato (par. 11 della sentenza). Sul piano storico, la sentenza ripercorre i lavori dell'Assemblea Costituente, notando che "[...] allorché si trattò di sostituire il termine "colpevole" con quello di "reo", dapprima si suggerì d'usare la parola "condannato" ma, successivamente, di fronte

In definitiva, mi pare che sostenere, oggi, una responsabilità penale singolare (e non personale, come ho tentato di spiegare) del sistema di IA significa tradire non solo la struttura dell'illecito penale, per via della rinuncia all'accertamento della colpevolezza, ma anche rinnegare la funzione rieducativa della pena, con l'effetto di rendere il diritto penale un diritto costituzionalmente sterile, poiché sprovvisto della finalità di rieducazione, e ispirato, piuttosto, a una logica di vendetta, ma anche questa prospettiva appare irragionevole se si considerano le difficoltà che si incontrano nel riconoscere la capacità del sistema di IA di soffrire<sup>11</sup>.

Ma procediamo con ordine. Prima di sostenere una responsabilità secondo questo itinerario del tutto innovativo che ci condurrebbe alle criticità di cui ho appena detto, bisogna percorrere l'itinerario più tradizionale e capire in che modo il diritto penale, secondo le sue logiche tradizionali e senza l'introduzione di nuovi centri di imputazione, risponda ai fenomeni di lesione o messa in pericolo di beni giuridici che si verificano per mezzo di un sistema di IA. In altri termini: fino a che punto si può spingere il rimedio penale 'classico' a fronte dei nuovi rischi collegati alla diffusione delle tecnologie di IA? E i problemi posti sono del tutto nuovi, come nuova è la tecnologia cui essi si riferiscono, oppure si tratta di questioni già in parte analizzate dal diritto penale del rischio?

A queste domande cercherò di dare una risposta, affrontando il primo di tre itinerari di responsabilità, rappresentati, rispettivamente, dalla responsabilità conseguente all'applicazione delle categorie tradizionali del diritto penale con riferimento al reato colposo, secondo la struttura tripartita dell'illecito; dalla responsabilità da reato degli enti

---

*alla contestazione sull'inusualità del termine "condannato" fuori dalla sede processuale, si tornò, per un momento, alla parola "colpevole", dichiarandosi espressamente: "Questa parola è più chiara, specialmente quando si parla di rieducazione del colpevole, perché il termine di rieducazione presuppone una colpa"* (par. 10 della sentenza).

<sup>11</sup> Allo stato attuale, pertanto, il destinatario della sanzione non può che essere il produttore e non il sistema di IA. E ciò vale anche per sanzioni diverse da quelle penali. Si può eventualmente pensare di prevedere delle sanzioni più mirate per i prodotti intelligenti. Ad esempio, la linea seguita dall'autorità americana garante della concorrenza (*Federal Trade Commission*) è quella di imporre la distruzione dell'algoritmo sviluppato in modo illecito. È quanto stabilito in ben tre decisioni tutte relative alla raccolta dei dati, la prima concernente il caso Cambridge Analytica del 2019, la seconda rivolta a un'applicazione di fotografie che aveva sfruttato la tecnologia di riconoscimento facciale, la terza, del 2022, di cui è stata destinataria l'azienda WW International, che aveva utilizzato dei dati raccolti in modo illecito in quanto concernenti dati di minori ottenuti senza il consenso dei genitori, la notizia è riportata dal sito computerworld.it in un articolo a firma di A. Grassi del 16.03.2022 dal titolo "Gli USA giustiziano l'algoritmo illegale: va cancellato" (<https://www.cwi.it/tecnologie-emergenti/intelligenza-artificiale/usa-giustiziano-algoritmo-illegale-144792>)

per effetto del meccanismo di cui all'art. 8 d.lgs. 231/2001; dalla responsabilità *singolare* del sistema di IA, cui ho già fatto brevemente cenno.



## SEZIONE I

### PECULIARITÀ DEL FATTO TIPICO COLPOSO NEL CONTESTO DELLE MACCHINE INTELLIGENTI

1. Il campo d'indagine: il reato colposo - 2. La dimensione oggettiva della colpa - 2.1 Genesi e individuazione della regola cautelare: la riconoscibilità e l'evitabilità - 3. Tipicità colposa e giudizio causale a confronto - 4. Principio di precauzione e cautele nelle attività sperimentali - 5. Rischio consentito - 6. La posizione del produttore. La perimetrazione della colpa specifica - 6.1 Produttore modello e prodotto modello - 6.2 Diligenza artificiale - 6.2.1 Base ontologica e base nomologica della riconoscibilità artificiale - 6.2.2. Negligenza artificiale - 7. La posizione dell'utente. Diligenza umana e affidamento - 7.1 Obblighi informativi dell'utente - 7.2 Dovere di vigilanza e dovere di intervento - 7.2.1 Macchine etiche? La violazione necessitata della regola di diligenza - 8. Diligenze a confronto

#### 1. Il campo d'indagine: il reato colposo

Nella trattazione delle categorie del diritto penale d'evento dinanzi ai nuovi rischi creati dalla produzione e dall'uso di sistemi intelligenti, il piano di indagine qui scelto è quello del reato colposo, che è anche il settore privilegiato di applicazione degli studi sul diritto penale del rischio<sup>1</sup>.

Le posizioni che vengono in rilievo, scartata l'idea di una responsabilità singolare del sistema intelligente, sono quelle del produttore e dell'utente-utilizzatore, il quale ultimo si atteggia non solo a vittima del reato colposo del produttore, ma anche ad autore di possibili fatti costituenti reato.

L'illecito colposo è tra le tipologie delittuose vissute “a lungo all'ombra del reato doloso”, considerato figura primordiale di illecito attorno al quale si è originariamente sviluppata la scienza penalistica. Come ben spiega Marinucci, la prassi e l'emersione di nuove esperienze criminologiche hanno alimentato la “liberazione” di categorie di reato diverse da quella (commissiva) dolosa. Così, dapprima fu prefigurata la divisione tra illeciti commissivi e omissivi e, successivamente, gli studi dottrinali tracciarono un sistema che valorizzava l'autonomia della colpa rispetto al dolo, favorendo la

---

<sup>1</sup> Ma è noto che le riflessioni sul rischio consentito o adeguato valgono anche per i reati dolosi, sul punto, V. Militello, *Rischio e responsabilità penale*, Milano, Giuffrè, 1988, p. 55 e pp. 205 ss.; M. Donini, voce *Imputazione oggettiva dell'evento (diritto penale)*, in *Enciclopedia del diritto, Annali, vol. III*, Milano, Giuffrè, 2010, pp. 646 ss.

particolarizzazione degli studi con riferimento agli illeciti commissivi dolosi e colposi e omissivi dolosi e colposi<sup>2</sup>. Di questa lenta emersione del reato colposo<sup>3</sup> v'è traccia all'interno del nostro codice, dove la colpa è descritta come un criterio d'imputazione soggettiva eccezionale rispetto al dolo, in quanto richiede sempre un'espressa previsione legislativa a mente dell'art. 43 c.p., sebbene dal punto di vista criminologico la colpa sia oggi un criterio di imputazione tutt'altro che eccezionale<sup>4</sup>; ma a ben vedere, già all'epoca del legislatore codicistico le occasioni di incidenti risultavano incrementate per via della diffusione massiccia del "macchinismo" che caratterizzava i processi produttivi e lo svolgimento delle mansioni quotidiane e iniziavano a emergere le esigenze di tutela tipiche di una 'società della tecnica'; la fenomenologia criminale prendeva allora le distanze da un modello legislativo in cui il reato colposo rappresenta l'eccezione.

La rinnovata vitalità del modello di imputazione colposa si affermerà definitivamente col passaggio alla società del rischio ed è confermata dal momento storico che stiamo vivendo, definito in dottrina come quarta rivoluzione<sup>5</sup>, in cui l'uomo non è più l'unico "organismo informazionale" ed è coadiuvato nelle sue attività da agenti artificiali intelligenti per lo svolgimento di certi compiti e mansioni.

In questo contesto, il tema del danno da prodotto conosce un nuovo prolungamento che spinge gli studiosi ad andare oltre la tralatizia "ottica umana" di approccio al tema e a "misurarsi con le prestazioni della *new entry AI*"<sup>6</sup>. Se le azioni in cui è coinvolto un sistema di IA diverranno penalmente rilevanti, lo saranno principalmente nella veste del reato colposo, commissivo e omissivo, sul quale la presente indagine si concentrerà. Del

---

<sup>2</sup> G. Marinucci, *Il reato come azione*, Milano, Giuffrè, 1971, pp. 132-133. Fu il Radbruch a prefigurare per primo la divisione del sistema in illeciti commissivi e omissivi (G. Radbruch, *Zur Systematik der Verbrechenlehre*, in *Festgabe für Frank*, 1930, ristampa anast. 1969, vol. I, p. 159).

<sup>3</sup> Sull'argomento cfr. D. Castronuovo, voce *Colpa penale*, *Enc. dir., I Tematici, Il reato colposo*, Milano, Giuffrè 2021, pp. 200ss., in particolare, p. 201-204.

<sup>4</sup> S. Canestrari, *La colpa*, in *Trattato di diritto penale, Parte generale, vol. II, Il Reato*, a cura di A. Cadoppi, S. Canestrari, A. Manna, M. Papa, Torino, UTET, 2013, p. 130.

<sup>5</sup> L. Floridi, *La quarta rivoluzione. Come l'infosfera sta cambiando il mondo*, Milano, Raffaello Cortina editore, 2017, in particolare pp. 99 ss. Il filosofo individua la prima rivoluzione nella scoperta dell'eliocentrismo da parte di Niccolò Copernico, con la pubblicazione del suo trattato *Sulle rivoluzioni dei corpi celesti* nel 1543, momento a partire dal quale l'uomo avrebbe cessato di credere nella sua centralità, quantomeno con riferimento al pianeta Terra. La seconda rivoluzione è fatta coincidere con la pubblicazione, nel 1859, de *L'origine delle specie* di Charles Darwin, che spinge l'uomo a rinunciare alla sua centralità nel mondo animale, dovendosi confrontare coll'idea che ogni specie vivente derivi da progenitori comuni attraverso un processo di selezione naturale. La terza rivoluzione giunge col lavoro psicoanalitico di Freud, il quale smentì l'idea della mente pensata alla stregua di una scatola conoscibile semplicemente guardandovi all'interno, dimostrando che molte nostre azioni sono frutto dell'inconscio. La quarta rivoluzione mette in discussione l'uomo quale unico essere insuperabile sul piano dell'intelligenza.

<sup>6</sup> C. Piergallini, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, in *Riv. it. dir. proc. pen.*, 4, 2020, pp. 1745-1774, in particolare p. 1749.

resto, le problematiche che susciterà un atteggiamento doloso dell'agente, sia esso il produttore o l'utente del sistema intelligente, non divergeranno dai temi che tradizionalmente sono stati affrontati nello studio del reato doloso<sup>7</sup>. Diverso è il caso del reato colposo, dove sorgono questioni relative all'individuazione delle regole oggettive di diligenza su base prasseologica ovvero su base giuridica, con particolare riferimento al ruolo che assumono a tal fine i criteri della prevedibilità e evitabilità dell'evento e la misura degli stessi alla luce della figura dell'agente modello; anche il riparto di responsabilità tra il produttore e l'utente appare problematico, soprattutto in presenza dei meccanismi di autoapprendimento dei sistemi di IA. Sullo sfondo resta il più esteso problema del bilanciamento tra lo svolgimento di un'attività pericolosa ma socialmente utile, qual è appunto la produzione di sistemi intelligenti, e la tutela degli interessi da essa minacciati, che ci spinge a interrogarci sull'estensione dell'area del rischio consentito.

Terreno d'elezione delle problematiche legate all'area del rischio consentito nelle attività produttive è quella responsabilità definita "per il tipo di produzione", contrapposta a quella "per il modo di produzione"<sup>8</sup>, di cui è bene dare brevemente conto prima di passare all'analisi delle questioni più specificamente legate alla responsabilità colposa, anche perché si tratta di nozioni che torneranno utili nel prosieguo.

L'espressione "tipo di produzione" allude alla specie di attività produttiva complessivamente considerata e la responsabilità per il tipo di produzione caratterizza i settori produttivi che presentano un grado di pericolosità talmente elevato (tanto all'interno, quanto all'esterno dello stabilimento) da essere difficilmente contrastabile mediante cautele relative al processo produttivo e tali da provocare danni gravi e irreversibili<sup>9</sup>. Su questo versante è emblematico lo svolgimento di attività al "alto rischio"<sup>10</sup>, in cui sul momento della responsabilità *ex post* del produttore, prevale a livello

---

<sup>7</sup> Riflette però sulla responsabilità a titolo doloso I. Salvadori, *Agenti artificiali, cit.*, p. 100, per concludere comunque che il caso non solleva particolari problemi; così anche R. Borsari, *Intelligenza Artificiale e responsabilità penale: prime considerazioni*, in *MediaLaws*, p. 264; a tal proposito, secondo F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. e uomo*, 29 settembre 2019, "Dobbiamo, insomma, prepararci ad un'era in cui la commissione di reati con lo strumento dell'IA potrebbe diventare assai frequente ed incisiva, anche in considerazione dell'accresciuta vulnerabilità di alcuni aspetti della vita umana connessi ad impieghi dell'intelligenza artificiale" e dunque si domanda, l'A.: "Occorre, allora, mettere in campo nuove fattispecie di reato (o rimodellare quelle già esistenti) al fine di renderle applicabili alla realizzazione di condotte criminose attraverso lo strumento dell'IA, offrendo così tutela ai beni giuridici anche da questa nuova fonte di attacchi?", pp. 26-27.

<sup>8</sup> F. Bricola, *Responsabilità penale per il tipo e il modo di produzione*, in *La responsabilità dell'impresa per i danni all'ambiente e ai consumatori*, Milano, Giuffrè, 1978, pp. 75 ss.

<sup>9</sup> *Ibidem*, p. 76.

<sup>10</sup> C. Piergallini, *Danno da prodotto e responsabilità penale*, Milano, Giuffrè, 2004, p. 41

legislativo la politica dei controlli amministrativi preventivi per autorizzare o negare determinati tipi di produzione<sup>11</sup>. La responsabilità per il tipo di produzione richiama le problematiche legate all'individuazione dell'area del rischio consentito, per delimitare le attività tollerate dall'ordinamento in ragione della loro utilità sociale pur nella prevedibilità che da esse possano derivare eventi dannosi<sup>12</sup>.

Il “modo di produzione” allude invece allo svolgimento dell'attività produttiva, che deve avvenire in condizioni di sicurezza ambientale sia interna all'azienda (con riguardo alle posizioni dei lavoratori), che esterna a essa<sup>13</sup>, vantando la collettività un interesse a che lo svolgimento della produzione avvenga secondo modalità che eliminino o riducano il rischio di danno<sup>14</sup>. In caso di responsabilità per il modo di produzione, il coefficiente di rischio è basso grazie all'esistenza di un patrimonio di regole cautelari in grado di neutralizzare il rischio, come avviene nella materia della prevenzione degli infortuni sul lavoro, che costituisce il principale terreno di elezione di questa forma di responsabilità<sup>15</sup>.

Con specifico riguardo alla responsabilità penale del produttore per i danni da prodotto, la dottrina ha osservato che si tratta di una responsabilità “trasversale”, che abbraccia tanto il tipo, quanto il modo di produzione<sup>16</sup>. Segnatamente, nella responsabilità per il tipo di produzione verrebbero in rilievo le qualità del prodotto, che si caratterizzerebbe per una pericolosità non integralmente contenibile né con misure cautelari ad hoc, né adottando determinate tecniche produttive. La responsabilità per le modalità di produzione evocherebbe invece la “funzionalizzazione del processo produttivo a eliminare o minimizzare il rischio di danno del prodotto”, il quale non presenterebbe profili di pericolosità residuale se costruito “a regola d'arte”<sup>17</sup>.

Volgendo l'attenzione più specificamente all'oggetto della produzione e ai suoi difetti, si suole distinguere tra difetti di costruzione, di fabbricazione, da informazione e da rischio di sviluppo<sup>18</sup>.

I primi, cioè i difetti di costruzione, sono imputabili all'errata progettazione o scelta dei materiali e si caratterizzano per colpire l'intera serie produttiva: il danno accompagna la vita del prodotto e ha una dimensione ubiquitaria. Diverso è il difetto di fabbricazione,

---

<sup>11</sup> F. Bricola, *cit.*, p. 76.

<sup>12</sup> C. Piergallini, *cit.*, p. 41.

<sup>13</sup> F. Bricola, *cit.*, p. 83.

<sup>14</sup> C. Piergallini, *Danno da prodotto, cit.*, p. 44.

<sup>15</sup> *Ibidem*, p. 44.

<sup>16</sup> *Ibidem*, p. 46.

<sup>17</sup> *Ibidem*, p. 46.

<sup>18</sup> È la distinzione elaborata da C. Piergallini, *op. cit.*, pp. 46 ss.

che colpisce solo alcuni elementi della serie ed estrinseca un rischio prevedibile in quanto statisticamente calcolabile. Il difetto da informazione è quello che avrebbe potuto essere evitato dotando il prodotto di istruzioni che ne descrivono in modo chiaro e completo l'impiego. Infine vi sono quei difetti imprevedibili al momento in cui il prodotto è immesso nel mercato, perché le conoscenze scientifiche disponibili non riescono a rintracciarli se non dopo che si sia verificato l'evento dannoso, sono questi i difetti da rischio da sviluppo.

È a questi difetti e i tipi di responsabilità individuati che si rivolgerà la nostra indagine sul fatto tipico colposo nel contesto delle macchine intelligenti, che a questo punto siamo pronti ad affrontare.

## 2. La dimensione oggettiva della colpa

Nell'ambito della concezione normativa della colpevolezza, tanto il dolo quanto la colpa caratterizzano una condotta dell'agente difforme dal contegno che doveva tenersi, ragione per la quale un rimprovero è possibile muovere all'autore del fatto doloso per aver tenuto un comportamento che non doveva volere, e a quello del fatto colposo, per aver realizzato un fatto involontario che non si doveva produrre<sup>19</sup>.

La colpa, però, più del dolo, si connota per il suo carattere normativo, che spinge alcuni autori a definirla come "elemento normativo" della fattispecie<sup>20</sup>, perché essa stessa è una nozione normativa, in quanto definibile solo in rapporto a una regola cautelare che preesiste<sup>21</sup>.

---

<sup>19</sup> G. Delitala, *Il "fatto" nella teoria generale del reato*, Padova, Cedam, 1930, p. 85; F. Mantovani, voce *Colpa*, in *Dig. Disc. Pen.*, 1988, vol. II, Padova, Cedam, pp. 299-314, in particolare p. 301.

<sup>20</sup> D. Castronuovo, *L'Evoluzione teorica della colpa penale tra dottrina e giurisprudenza*, in *Riv. it. dir. proc. pen.*, 2011, pp. 1594-1645, in particolare, p. 1616.

<sup>21</sup> S. Canestrari, *La colpa*, cit., p. 137. Va però accolta la precisazione di quella dottrina che osserva come "le cautele non sono elementi normativi del fatto" perché se così fosse, l'errore sulla regola cautelare dovrebbe risolversi in un errore sul fatto con conseguente applicazione dell'art. 47 c.p. Ciò può al più verificarsi quando la cautela è contenuta in una legge penale o è incorporata in essa per rinvio, casi, questi, in cui si sarà sì in presenza di un errore su legge penale. Così M. Donini, *L'elemento soggettivo della colpa. Garanzie e sistematica*, in *Riv. it. dir. proc. pen.*, 2013, pp. 124 ss., in particolare p. 135 e 137. Tuttavia, da tali osservazioni, cioè dal fatto che "le cautele non sono elementi normativi del fatto", l'autore trae la conclusione che "la misura oggettiva della colpa non esiste", nel senso che "la colposità non è un connotato intrinseco (*in re ipsa*) della condotta (a sua volta oggetto di dolo o colpa), se non in relazione a un soggetto e a una situazione concreti nella quale si esprime", p. 137. L'A. non critica il ruolo riconosciuto alla misura oggettiva nell'accertamento del fatto tipico colposo (anzi riconosce che si tratta di un "momento di contestazione processuale di una condotta generalizzabile ai fini di garanzia della colpa concreta", p. 138 che "attiene alle garanzie processuali della contestazione di un fatto colposo esigibile *ex ante*" p. 139), cioè non ripudia la teoria della doppia misura, ma piuttosto critica la locuzione "misura oggettiva della colpa", poiché "una colpa oggettiva è come un dolo oggettivo: è una *culpa in re ipsa*, (...) è l'inizio di un discorso

Tre sono allora gli elementi caratteristici della colpa: 1) l'elemento negativo della mancanza di volontà del fatto materiale tipico; 2) l'elemento oggettivo della inosservanza di regole di condotta dirette a prevenire danni ai beni protetti; 3) l'elemento soggettivo dell'attribuibilità di tale inosservanza al soggetto agente, la cui capacità di adeguarsi alle regole fonda la pretesa giuridica del loro rispetto<sup>22</sup>.

La distinzione tra elemento oggettivo e elemento soggettivo della colpa fonda l'idea di una 'misura doppia' della stessa, in cui si rispecchierebbero rispettivamente l'illiceità e la colpevolezza del delitto colposo<sup>23</sup>. L'accertamento del primo momento della colpa attiene al piano oggettivo, relativo al "dovere impersonale di agire", cui segue la verifica della colpevolezza colposa, cioè la rimproverabilità del fatto colposo all'agente in ragione di un "potere personale di agire"<sup>24</sup>, che rende la verifica della trasgressione della regola cautelare elemento necessario -benché non sufficiente- per l'identificazione di un reato colposo<sup>25</sup>.

In questa Sezione I ci occuperemo del fatto tipico colposo, per individuare alcune sue peculiarità nel contesto delle macchine intelligenti. A tal fine, bisogna tracciare, seppur brevemente, le linee essenziali della normatività della colpa, che torneranno utili per avviare una riflessione sui doveri di diligenza che gravano sul produttore e sull'utente di un sistema di IA.

Restando allora sul piano di indagine individuato, cioè quello della dimensione oggettiva della colpa, prima di addentrarci sui problemi attinenti alla genesi e all'individuazione delle regole cautelari, va brevemente evidenziato il ruolo che l'evento ha nel definire un comportamento come colposo.

Una certa condotta è colposa non solo se è accertata la riconoscibilità dell'evento sulla base delle conoscenze della figura modello; va infatti anche verificato che l'evento che rientri tra quelli che possono essere evitati astenendosi dalla condotta antidoverosa, cioè appartenga alla classe di rischi che la regola cautelare è tesa a contrastare<sup>26</sup>. L'evento

---

incostituzionale", p. 138. Si tratta di un'osservazione linguistica pregevole, ma che non è stata seguita dalla prevalente dottrina che ormai abitualmente impiega i termini "misura oggettiva" e "soggettiva" della colpa..

<sup>22</sup> Così, F. Mantovani, *voce Colpa, cit.*, p. 303.

<sup>23</sup> Sul punto, si rinvia a F. Giunta, *Illiceità e colpevolezza nella responsabilità colposa. I - La Fattispecie*; Padova, Cedam, 1993.

<sup>24</sup> S. Canestrari, *La colpa, cit.*, p. 143.

<sup>25</sup> S. Canestrari, *La colpa, cit.*, p. 137-138.

<sup>26</sup> M. Gallo, *voce, cit.*, p. 640; G. Marinucci, *La colpa, cit.*, p. 113 ss.; G. Forti, *cit.*, p. 359 ss.; P. Veneziani, *Regole cautelari proprie e improprie nella prospettiva delle fattispecie colpose causalmente orientate*, Padova, Cedam, 2003, pp. 49 ss.

ha quindi un ruolo nella selezione della condotta tipica, per cui la cautela scritta deve contrastare un evento che presenti certi caratteri essenziali, in carenza dei quali il solo disallineamento da quanto in essa previsto non è sufficiente a fondare il rimprovero per colpa. In altri termini, l'evento realizzatosi deve concretizzare il rischio tipico che la regola cautelare mirava a prevenire, non potendosi rispondere per la violazione della stessa allorché l'evento materializzi un rischio estrinseco, non evitabile neppure da parte di una condotta conforme alla regola di diligenza, cioè mediante l'astensione dall'azione, da certe modalità dell'azione o adottando determinate cautele<sup>27</sup>. Questo ragionamento porta inevitabilmente a riconoscere che l'evento appartiene alla struttura dei reati colposi causalmente orientati, rappresentandone "l'estremo"<sup>28</sup>. Se così non fosse, se cioè si estromettesse l'evento dalla struttura del reato colposo causalmente orientato, si annullerebbe la distinzione tra evento cagionato *per colpa* (art. 43, comma 3) e evento posto *altrimenti* a carico dell'agente come conseguenza della sua azione o omissione (art. 42, comma 3)<sup>29</sup>. Invece è proprio il dato positivo a suggerire che l'evento nei reati colposi causalmente orientati non può essere addossato all'agente sulla base del solo nesso di causalità, come emerge dal fatto che l'evento deve essersi verificato *a causa* non di una condotta qualsiasi, ma di quella negligente, imprudente, imperita o inosservante di regole cautelari positivizzate. Non basta allora il solo nesso di causalità, ritenuto invece sufficiente nei casi di responsabilità oggettiva: l'evento deve essere "immancabilmente abbracciato da negligenza, imprudenza, imperizia"<sup>30</sup>.

Da qui la *doppia tipicità dell'evento* nei reati colposi causalmente orientati: esso dovrà essere anzitutto *tipico* rispetto al rischio coperto dalla regola cautelare e, in secondo luogo, *tipico* rispetto alla fattispecie incriminatrice<sup>31</sup>.

---

<sup>27</sup> M. Gallo, voce *Colpa penale (dir. vig.)*, in *Enc. dir.*, Milano, Giuffrè, 1960, p. 640.

<sup>28</sup> G. Marinucci, *La colpa, cit.*, p. 116.

<sup>29</sup> M. Gallo, voce, *cit.*, p. 634.

<sup>30</sup> G. Marinucci, *La colpa, cit.*, p. 138. L'A. invoca a conferma di questa lettura, oltre al dato letterale dell'art. 43 c. 3 e la distinzione, già sottolineata dal Gallo, con l'art. 42, c. 3, anche la categoria dei reati aggravati dall'evento, nei quali il reato base può essere costituito tanto da una fattispecie dolosa, che colposa. Ciò dimostrerebbe che se il reato colposo può costituire il reato base cui segue un evento aggravatore, allora nel primo l'evento non potrà essere imputato a titolo meramente oggettivo, cioè sulla base del solo rapporto causale, come invece accade per l'evento aggravatore che ricade nella logica del *versari in re illicita* (pp. 139-140).

<sup>31</sup> C. Piergallini, *cit.*, p. 406.

## 2.1 Genesi e individuazione della regola cautelare: la riconoscibilità e l'evitabilità

Si è detto come la regola cautelare sia parte integrante della fattispecie colposa, che si caratterizza per la violazione della stessa. Da questa incidenza della regola prudenziale sulla fattispecie penale è possibile trarre la definizione di regole cautelari: esse sono norme modali (quindi comportamentali) che indicano il modo in cui un'attività pericolosa, ma lecita, deve essere svolta, conciliando il diritto di libertà al loro svolgimento con la tutela di terzi<sup>32</sup>.

La regola cautelare può essere utilizzata nell'accertamento della tipicità colposa a condizione che sia già venuta ad esistenza, cioè sia disponibile e conoscibile alla figura modello<sup>33</sup>, da qui la distinzione tra *genesi* e *individuazione* della regola cautelare<sup>34</sup>. Il *meccanismo genetico* delle regole cautelari è fondato sui criteri della riconoscibilità e evitabilità dei rischi insiti nello svolgimento di una certa attività alla luce di giudizi ripetuti nel tempo e cristallizzatisi e, solo nel caso di colpa specifica, positivizzati dall'autorità o dal soggetto che pone la norma; l'*individuazione* delle cautele è invece fondata sul parametro di accertamento dell'agente modello<sup>35</sup>. La distinzione tra *genesi* e *individuazione* chiarisce che la regola cautelare deve necessariamente preesistere alla commissione del fatto e alla formulazione dell'imputazione<sup>36</sup>. La tipicità colposa, quindi, è accertata attraverso un giudizio di riconoscibilità del rischio e evitabilità dell'evento effettuato *ex ante*, cioè al momento in cui è realizzata la condotta, in base al parametro dell'agente modello<sup>37</sup>.

Risulta così chiarito che i criteri tesi a individuare la regola che l'agente deve osservare nello svolgimento di una data azione sono quelli della *rappresentabilità* (o *riconoscibilità*) e della *prevenibilità* (o *evitabilità*) dell'evento<sup>38</sup>: la tipicità colposa è costituita dalla violazione della diligenza, cioè dal non essersi resi conto di un pericolo

---

<sup>32</sup> Così, F. Giunta, *La normatività della colpa penale. lineamenti di una teorica*, in *Riv. it. dir. proc. pen.*, 1999, pp. 86- 115, in particolare p. 88.

<sup>33</sup> D. Castronuovo, *L'evoluzione teorica della colpa penale tra dottrina e giurisprudenza*, in *Riv. it. dir. proc. pen.* 2011, p. 1626.

<sup>34</sup> *Ibidem*; S. Canestrari, *La doppia misura della colpa*, in *Ind.pen.* 2005, p. 21 e 22; S. Canestrari, *La colpa*, *cit.*, p.144.

<sup>35</sup> S. Canestrari, *La doppia misura*, *cit.*, pp. 21 e 22.

<sup>36</sup> D. Castronuovo, *L'evoluzione teorica della colpa*, *cit.*, p. 1627.

<sup>37</sup> S. Canestrari, *La doppia misura*, *cit.*, p. 22.

<sup>38</sup> M. Gallo, *cit.*, p. 638, il quale precisa di prediligere il termine *rappresentabilità* anziché "prevedibilità" in ragione della maggiore ampiezza del primo, in grado di abbracciare non solo accadimenti futuri, ma anche quelli concomitanti o antecedenti all'azione del soggetto (nota 38).

che si aveva il dovere di riconoscere e dal non aver impiegato i mezzi necessari a evitare la verifica del fatto<sup>39</sup>. Questi due criteri della rappresentabilità e evitabilità rilevano su un duplice livello: in prima battuta, nella determinazione delle regole disciplinanti una certa attività e, successivamente, in sede di accertamento concreto della colpa<sup>40</sup> attraverso il parametro dell'agente modello.

A questo punto bisogna fare due precisazioni. La prima è che il giudizio di evitabilità del fatto va svolto a valle del giudizio di riconoscibilità: l'agente avrebbe dovuto evitare il fatto allorché la sua verifica era riconoscibile. L'evitabilità è quindi agganciata alla riconoscibilità.

La seconda è che tanto la riconoscibilità quanto l'evitabilità hanno in sede di *Tatbestand* una connotazione normativa, cioè vanno lette come *dovere* di riconoscere e *dovere* di evitare, se si aderisce all'impostazione, qui condivisa, per la quale la tipicità colposa rientra nella sfera del *Sollen* e non del *Können*<sup>41</sup>. Ne deriva che la possibilità di riconoscere, e quindi di evitare, l'accadimento vietato va valutata non in relazione a ciò che l'agente concreto avrebbe *potuto* fare, bensì a quanto avrebbe *dovuto* fare in base al parametro dell'agente modello. La norma cautelare è, quindi, regola modale che esprime l'evitabilità dell'evento in termini impersonali, ancora insufficiente a fondare un giudizio di colpevolezza del singolo agente<sup>42</sup>. È per questo che all'interno della categoria della riconoscibilità la dottrina distingue tra la possibilità "individuale" di riconoscere il verificarsi dell'evento e la possibilità "generale" di riconoscere<sup>43</sup>: la connotazione normativa del *Tatbestand* rende estranea a quest'ultimo la capacità individuale dell'agente, che al più rileverà in sede di colpevolezza<sup>44</sup>; sono quindi in questa fase irrilevanti stati e condizioni personali che hanno in concreto impedito la rappresentazione

---

<sup>39</sup> G. Forti, *cit.*, p. 205. L'A. precisa anche che (pp. 189 ss.) il canone delle evitabilità è pure utilizzato nella dottrina tedesca per definire un concetto unitario di azione, analogamente a quanto avviene per opera di una parte della dottrina italiana nell'interpretare il parametro della "coscienza e volontà" richiesto dall'art. 42, c.1 c.p., contrapponendosi all'opinione di chi invece ritiene che l'evitabilità (*rectius* la coscienza e la volontà) esaurirebbe la propria rilevanza nell'ambito delle sole fattispecie colpose. Nota l'A., che entrambe le opinioni mostrano l'indiscusso ruolo preminente del principio dell'evitabilità. La questione sarà oggetto di una più approfondita analisi più avanti.

<sup>40</sup> M. Gallo, *cit.*, p. 638.

<sup>41</sup> Questo è il riflesso della concezione normativa della colpa, contrapposta a una visione meramente psicologica della stessa. Sulle due concezioni, si rinvia a F. Mantovani, *voce Colpa*, in *Dig. Disc. Pen.*, 1988, vol. II, Padova, Cedam, pp. 299-314; G. Marinucci, *La colpa*, *cit.*, pp. 145 ss.; F. Giunta, *Illiceità e colpevolezza*, *cit.*, cap. I, pp. 4 ss.; G. Marini, *voce Colpevolezza*, in *Dig. Disc. Pen.*, 1988, vol. II, Padova, Cedam, pp. 314-322.

<sup>42</sup> F. Giunta, *La normatività*, *cit.*, p. 90.

<sup>43</sup> G. Forti, *cit.*, p. 205.

<sup>44</sup> Salvo quanto si preciserà di seguito con riguardo all'individualizzazione dello standard di diligenza operata mediante la figura dell'agente modello.

del rischio<sup>45</sup>. Pertanto, è la possibilità di riconoscere “in generale” la verifica dell’evento a caratterizzare quel ‘dovere di riconoscere’ tipico del fatto colposo che giustifica il rimprovero penale.

### 3. Tipicità colposa e giudizio causale a confronto

In presenza di quali condizioni la riconoscibilità del verificarsi del fatto può dirsi doverosa in quanto riferibile in via generale a un numero indefinito di individui? La dottrina risponde al quesito partendo dal dato per cui la riconoscibilità doverosa “è tanto maggiore quanto più alta è la probabilità oggettiva di verifica del fatto”, cioè quanto è più probabile che da una data azione segua un dato evento alla luce di generalizzazioni o leggi causali<sup>46</sup>. Asserire ciò significa ammettere che alla base del giudizio di riconoscibilità doverosa vi sono le stesse leggi causali utilizzate per stabilire il nesso di causalità<sup>47</sup>.

Ciononostante, resta fermo che il giudizio è diverso da quello causale, in ragione del fatto che il giudizio di riconoscibilità generale stabilisce un nesso di adeguatezza causale tra la condotta e l’evento non *ex post*, come avviene in sede di causalità, ma *ex ante*, essendo valutato in astratto, cioè prima che si verifichi una situazione concreta<sup>48</sup>. L’astrattezza non consente, in questa fase dell’elaborazione della nozione di riconoscibilità, di prendere in esame la totalità degli antecedenti concreti di cui invece non si può prescindere in sede di causalità, ove le leggi di copertura servono a risalire alle condizioni iniziali a partire da un fatto già verificatosi. La prospettiva *ex ante* in cui ci pone la riconoscibilità del pericolo garantisce che la cautela preesiste, deve cioè essere individuata prima che il danno si verifichi, quindi “il cittadino e il giudice non sono facitori, bensì fruitori di cautele, al pari delle leggi causali”<sup>49</sup>. Pertanto, è suscettibile di individuazione, tanto da parte del cittadino, quanto del giudice, solo la regola di comportamento già venuta ad esistenza e già disponibile, da qui la tendenza in dottrina a distinguere tra genesi e individuazione delle regole cautelari per affermare che il processo

---

<sup>45</sup> M. Gallo, *cit.*, p. 640.

<sup>46</sup> G. Forti, *cit.*, p. 208.

<sup>47</sup> G. Forti, *cit.*, p. 208.

<sup>48</sup> *Ibidem*, p. 208 e anche p. 230.

<sup>49</sup> F. Giunta, *Culpa, culpa*, in *disCrimen*, 4.06.2019, p. 19.

genetico deve necessariamente precedere il momento dell'individuazione<sup>50</sup>, che appartiene invece al procedimento interpretativo di concretizzazione della colpa<sup>51</sup>.

Le leggi causali rappresentano allora uno strumento utile al giudice per l'accertamento del nesso di causalità, ma, ancor prima, utile all'ordinamento nel processo genetico della cautela per stabilire il dovere di riconoscere rilevante ai fini della tipicità colposa, che presenta la medesima struttura logica di un giudizio probabilistico. A questo livello, quindi, i requisiti della prevedibilità e dell'evitabilità<sup>52</sup> operano ancora solamente nel procedimento genetico delle regole cautelari (positivizzate e non), in una prospettiva *ex ante* che mostra il risultato tipico come una conseguenza probabile o ragionevolmente possibile di una data condotta<sup>53</sup>; sempre la rappresentabilità e l'evitabilità dell'evento tipico rileveranno poi anche in una prospettiva *ex post*, nel momento dell'accertamento della causalità nei reati d'evento, in applicazione della formula della *condicio sine qua non*<sup>54</sup>.

Restando sempre sul piano della tipicità colposa, va osservato come secondo la dottrina le leggi causali compongono uno dei due fattori della riconoscibilità e, in particolare, la sua componente nomologica, cui si aggiunge il secondo fattore che è rappresentato dalla componente ontologica della riconoscibilità<sup>55</sup>. Quest'ultima comprende tutti gli elementi fattuali in grado di connettere le conoscenze nomologiche alla situazione concreta, nel senso che è la presenza di certe condizioni iniziali che consente all'agente di leggere la situazione concreta alla luce delle leggi causali e quindi di poter riconoscere la verificazione dell'evento come uno sviluppo della condotta contraria al dovere di diligenza. In altri termini, sono gli elementi ontologici ad "attualizzare" quanto stabilito dalle conoscenze nomologiche.

Leggendo la componente nomologica e quella ontologica alla luce della natura normativa della riconoscibilità (cioè riconoscibilità non come "possibilità di riconoscere", bensì come "dovere di riconoscere" o *Erkennensollen*) di cui si è sopra detto, risulta che la riconoscibilità consiste nel *dovere* di possedere un determinato *insieme di conoscenze relative a generalizzazioni* e nel *dovere* di percepire la presenza di elementi fattuali che,

---

<sup>50</sup> D. Castronuovo, *L'evoluzione teorica*, cit., p. 1626

<sup>51</sup> *Ibidem*, p. 1627.

<sup>52</sup> Sulla prevedibilità, si veda il contributo di G. Civello, voce *Prevedibilità e reato colposo*; sull'evitabilità, K. Summerner, voce *Evitabilità dell'evento e comportamento alternativo lecito*, entrambi in in *Enc Dir., I Tematici, Il reato colposo*, Milano, Giuffrè, rispettivamente a p. 1004 ss. e p. 489 ss.

<sup>53</sup> V. De Francesco, *Sulla misura soggettiva della colpa*, in *Studi Urbinati*, v. 30, 1978, p. 278.

<sup>54</sup> M. Gallo, voce, cit., pp. 638-639, nota n. 39, sebbene si riferisca alla sola evitabilità.

<sup>55</sup> G. Forti, cit., pp. 216 ss.

combinata alle conoscenze nomologiche, consentono di “anticipare” la verifica di possibili lesioni a beni giuridici<sup>56</sup>.

Anche qui va evidenziata una differenza rispetto al giudizio di causalità materiale: in sede di accertamento della misura oggettiva della colpa si terrà conto delle sole conoscenze nomologiche disponibili al momento del fatto e non anche di quelle scoperte in epoca successiva. Qui il giudizio ha comunque una base soggettiva, costituita dalle conoscenze possedute dall’agente modello secondo la logica della valutazione *ex ante*, cosicché l’accertamento dell’agire colposo è privato dell’oggettività tipica del giudizio causale, il quale è invece formulato sulla base di tutte le conoscenze sociali e scientifiche disponibili al momento del fatto<sup>57</sup>. Questa è la ragione per la quale se, da un lato, è giusto sottolineare il connotato normativo della riconoscibilità e della evitabilità quali elementi del fatto colposo, dall’altro va però dato atto che non è del tutto vero che non rilevino in alcun modo le caratteristiche dell’agente concreto in sede di tipicità colposa. Piuttosto, va detto che esse acquistano valore nei limiti in cui siano oggettivate nella *Maßfigur*, cioè in quel soggetto immaginato come personificazione dell’ordinamento giuridico nella situazione concreta. L’agente modello costituisce “il parametro attraverso il quale una sfera del “dovere” viene ritagliata dalla sfera del “potere””, poiché “il dovere di riconoscere è la possibilità di riconoscere rapportata al punto di vista dell’agente modello”<sup>58</sup>. L’effetto è quello di individualizzare il livello di diligenza alla luce del parametro dell’*homo eiusdem professionis et condicionis*, selezionando le conoscenze che l’ordinamento pretende che l’appartenente a una data categoria sociale e professionale possieda per intraprendere una certa attività.

Tutto ciò porta a ritenere come i giudizi di riconoscibilità ed evitabilità abbiano comunque una base soggettiva che, se non può essere ricondotta allo stesso soggetto agente in concreto<sup>59</sup>, è però costituita dalle conoscenze possedute dalla figura modello. La *Maßfigur* si colloca nel punto di equilibrio tra due esigenze contrapposte, cioè quella di non soggettivizzare eccessivamente il giudizio di colpa fino a renderlo inattuabile e

---

<sup>56</sup> *Ibidem*, p. 219.

<sup>57</sup> G. Forti, *cit.*, p. 211 ss.; C. Piergallini, *cit.*, p. 406; G. Marinucci, *Innovazioni tecnologiche e scoperte scientifiche: costi e tempi di adeguamento delle regole di diligenza*, in *Riv. it. dir. proc. pen.*, 2005, I, pp. 30-60, in particolare p. 52.

<sup>58</sup> G. Forti, *cit.*, p. 234.

<sup>59</sup> Se così fosse, infatti, si ‘svuoterebbero’ la rappresentabilità e la evitabilità della loro funzione di criteri in grado di individuare un agire cui conformare il proprio, perché nella prospettiva dell’agente concreto l’evento sarà sempre non riconoscibile e quindi non prevenibile, M. Gallo, *voce, cit.*, p. 639.

quella di impregnare la condotta diligente di almeno alcune delle caratteristiche della situazione concreta<sup>60</sup>.

È, questo, un ulteriore punto che distingue l'accertamento della tipicità colposa dal nesso di causalità: se in entrambi i casi impieghiamo leggi o generalizzazioni che delineano un nesso di derivazione tra una data condotta e un certo evento, solo nel giudizio causale si accerta la probabilità oggettiva di verifica del fatto, che, per contro, in sede di accertamento del fatto colposo sarà attribuito alla negligenza dell'agente solo se la sua realizzazione era riconoscibile all'*homo eiusdem professionis et condicionis* alla luce delle conoscenze che era legittimo presumere in lui alla luce della posizione sociale rivestita e dell'attività svolta<sup>61</sup>.

Ne deriva che le condotte contrarie a diligenza costituiscono un sottoinsieme delle condotte che, sulla base di leggi scientifiche causali, si ritiene inneschino un certo rischio<sup>62</sup>.

Ci si può a questo punto domandare quale debba essere il livello di certezza con la quale la scienza faccia risalire un certo fatto a una data condotta, quindi se solo leggi causali che offrono un certo livello di certezza vadano impiegate nel giudizio di probabilità e se la risposta debba essere la stessa sia sul piano della riconoscibilità quale elemento della colpa, che sul piano della causalità.

La questione ha rappresentato uno dei temi più caldi attinenti all'accertamento del nesso di causalità, per il quale la giurisprudenza ha proposto la soluzione di compromesso della probabilità logica, come si avrà modo di approfondire più avanti.

Rimandando quindi a altra sede la trattazione del livello di certezza delle leggi scientifiche ai fini del giudizio causale, è però qui doveroso concentrarci su quali leggi causali siano utilizzabili nel procedimento genetico della regola cautelare.

#### **4. Principio di precauzione e cautele nelle attività sperimentali**

Nel contesto delle tecnologie di IA deve ribadirsi che l'incriminazione delle condotte non può passare attraverso una cautela desunta da conoscenze non certe, che in sostanza introdurrebbero il principio di precauzione tra i criteri di incriminazione. Una simile

---

<sup>60</sup> M. Gallo, *voce, cit.*, p. 639

<sup>61</sup> M. Gallo, *cit.*, p. 640, il quale precisa che sono invece irrilevanti stati e condizioni personali che hanno in concreto impedito la rappresentazione del rischio.

<sup>62</sup> G. Forti, *cit.*, p. 250.

affermazione è il precipitato della stessa struttura normativa della colpa sopra analizzata, la cui natura nega un ruolo al principio di precauzione in materia penale<sup>63</sup>.

Si è visto infatti come il rimprovero per colpa presuppone l'esistenza del rischio e la riconoscibilità e evitabilità dello stesso da parte dell'agente modello. Nella colpa per violazione di una regola desunta dal principio di precauzione questi requisiti mancherebbero: l'esistenza del rischio è incerta e, di conseguenza, non è possibile individuare una condotta adeguata a evitarlo<sup>64</sup>. Tuttavia è la stessa struttura del rimprovero per colpa a richiedere che la norma cautelare sia posta per evitare rischi che sono già generalmente prevedibili e non anche quelli che lo diventano solo dopo che l'offesa si sia realizzata; a ritenere diversamente, la regola di condotta sarebbe formulata "a posteriori"<sup>65</sup> e a farne le spese sarebbe la libertà personale di chi quel rischio non poteva in alcun modo riconoscere.

La condotta stessa dell'agente modello ne risulterebbe modificata: il rischio incerto che trascinerebbe il principio di precauzione all'interno della norma incriminatrice porrebbe l'agente dinanzi l'alternativa tra l'agire e il non agire, mentre non esisterebbe un agire conforme a diligenza in grado di eliminare il rischio di verificazione dell'evento, poiché l'esistenza stessa del rischio non è stabilita. L'opzione più drastica cui porterebbe il "precauzionismo" sarebbe allora quella di fronteggiare il pericolo insito nella produzione di sistemi intelligenti attraverso la regola dell'astensione<sup>66</sup>; nondimeno si tratta di una strada non percorribile per gli indubbi vantaggi che dal progresso tecnologico derivano<sup>67</sup>. Inoltre, con riferimenti alla produzione di sistemi intelligenti -e non solo- va

---

<sup>63</sup> Sul principio di precauzione, F. D'Alessandro, voce *Precauzione (principio di)*, *Enc. dir., I Tematici, Il reato colposo*, Milano, Giuffrè, 2021, p. 975 ss.; F. Giunta, *Il diritto penale e le suggestioni del principio di precauzione*, *Criminalia*, 2006, pp. 227 ss.; C. Ruga Riva, *Principio di precauzione e diritto penale. Genesis e contenuto della colpa in contesti di incertezza scientifica*. In E. Dolcini, & C.E. Paliero (a cura di), *Studi in onore di Giorgio Marinucci (pp. 1743-1777)*, Milano, Giuffrè, 2006; A. Orsina, *Rischio da incertezza scientifica e modelli di tutela penale*, Giappichelli, Torino, 2015.

<sup>64</sup> E. Dreyer, *Droit pénal et principe de précaution*, in *Recueil Dalloz*, 1 oct. 2015, n. 33, pp. 1912-1918.

<sup>65</sup> L. Stortoni, *Angoscia tecnologica ed esorcismo penale*, in *Riv. it. dir. proc. pen.*, 2004, p. 80.

<sup>66</sup> Sono qui riadattate al contesto della produzione dei sistemi di IA le espressioni utilizzate da F. Giunta, *Il diritto penale e le suggestioni del principio di precauzione*, in *Criminalia*, 2006, p. 227 ss., in particolare p. 232.

<sup>67</sup> I. Salvadori, *Agenti artificiali*, cit., p. 115; secondo M. Di Florio, *Il diritto penale che verrà. Brevi considerazioni sul possibile impiego dell'IA per prevenire il rischio di disastri colposi*, in *Arch. Pen.*, 2, 2021 l'IA potrebbe essere impiegata per verificare la sicurezza di infrastrutture stradali al fine di ridurre i rischi di pericoli alla sicurezza e all'incolumità degli utenti, p. 12; v'è poi chi valorizza l'apporto dell'IA nel processo di standardizzazione dei modelli organizzativi richiesti dal d.lgs. 231/2001, così R. Trezza, *L'Intelligenza Artificiale come ausilio alla standardizzazione del modello 231: vantaggi "possibili" e rischi "celati"*, in *Giur. pen.*, 2021/1-Bis; più in generale sull'uso di algoritmi per una protezione più efficace e efficiente dei beni giuridici, si rinvia alle riflessioni di C. Burchard, *L'intelligenza artificiale*

considerato che una scelta di tal tipo rischierebbe di ingessare lo sviluppo di un intero settore economico all'interno dell'ordinamento che vieta la produzione, senza però con ciò eliminare il rischio, che rimarrebbe attuale per via della delocalizzazione della produzione di sistemi di IA in quei paesi i cui ordinamenti mostrano maggiori aperture verso il progresso.

Lasciando il terreno delle ragioni legate allo sviluppo tecnologico e di tipo economico, e ritornando alle resistenze poste a livello ordinamentale, va ancora evidenziato come ostacoli al riconoscimento di un'efficacia incriminatrice al principio di precauzione che si otterrebbe traducendo il "precauzionismo" in regola di condotta si incontrano anche sul piano dei principi che interessano l'intero diritto penale e vanno oltre la struttura del fatto tipico colposo, che già, come si è detto, rappresenta un impedimento per principio in esame. Segnatamente, punendo un comportamento che genera un pericolo solo potenziale, individuato sulla base di una regola scientifica incerta, si aprirebbe uno scontro aperto col principio di offensività a causa della genesi di un diritto penale eccessivamente e irragionevolmente sbilanciato verso l'interesse di vittime solo potenziali (e non attuali) a discapito della libertà dell'agente, così rovesciando i valori di una società democratica, in cui la libertà non è (e mai può considerarsi) secondaria a una avvertita esigenza di sicurezza contro un pericolo che è addirittura incerto<sup>68</sup>. Risulterebbe in altre parole negato il primato della funzione garantistica della colpevolezza, in favore di valutazioni che attengono al bisogno di pena rese possibili dall'arretramento della punibilità che il principio di precauzione comporterebbe se riuscisse a penetrare nella struttura del reato<sup>69</sup>. Trasponendo queste conclusioni nel campo del reato colposo, diremmo che l'incertezza scientifica farebbe assumere alla colpa "contorni di una genericità inammissibile" e aprirebbe la strada verso forme di responsabilità obbiettiva, perché si ometterebbe di precisare la cautela che in concreto andrebbe adottata o il comportamento da evitare<sup>70</sup>. Sono allora le elementari e fondamentali regole di giustizia a imporre la riconoscibilità ex ante della norma di comportamento e la ragionevole prevedibilità della valutazione giudiziale<sup>71</sup>.

---

*come fine del diritto penale? sulla trasformazione algoritmica della società*, in *Riv. it. dir. proc. pen.*, 4, 2019, pp. 1909 ss.

<sup>68</sup> E. Dreyer, *op. cit.*, p. 1917.

<sup>69</sup> F. Giunta, *Il diritto penale e le suggestioni*, *cit.*, p. 229.

<sup>70</sup> M. Gallo, *voce*, *cit.*, p. 638.

<sup>71</sup> Così, F. Giunta, *Il diritto penale e le suggestioni*, *cit.*, p. 229.

Che il problema dell'impiego di conoscenze tecnico-scientifiche incerte possa presentarsi nella materia dell'intelligenza artificiale è fuor di dubbio, considerato anche il fatto che molte tecnologie, sebbene note da tempo (è del 1956 la conferenza tenutasi al Dartmouth College), sono rimaste spesso in una fase di sperimentazione anche a causa di quei periodi di sfiducia verso questo settore della scienza che ha comportato una diminuzione degli investimenti nel settore. Solo negli ultimi decenni questa branca scientifica ha riacquisito la fiducia degli investitori, dopo che la ricerca è stata indirizzata verso le forme di IA 'debole' o 'leggera', per abbandonare l'illusione della creazione di meccanismi artificiali che riproducessero le funzioni e le capacità del cervello umano, con ciò svuotando di significato l'espressione 'intelligenza artificiale', trattandosi più di una potenziata capacità di calcolo eseguita anche senza supervisione umana<sup>72</sup>.

Per non affidare l'imputazione penale all'incertezza scientifica sarà sempre necessaria una importante attività di sperimentazione dei sistemi prima della loro immissione nel mercato. In quest'ottica, un momento importante nella fase di produzione dell'algoritmo di autoapprendimento è quella della validazione dei risultati raggiunti in fase di test.

Un problema sorge però nel caso in cui già la sperimentazione esponga a rischi beni presidiati dalla norma penale. Questo scenario si verifica ad esempio quando le varie componenti tecnologiche del sistema sono assemblate e vengono testate in contesti che simulano quelli reali in cui andranno a operare. È quanto accade allorché un'auto con pilota automatico sarà testata in condizioni che simulano il traffico delle città, che concretizzano un rischio per i beni dell'integrità fisica e della vita. Qui la dinamica fattuale si svolge entro il confine dell'incertezza scientifica propria delle sperimentazioni, non assistite da regole cautelari specifiche e consolidate<sup>73</sup>, per cui una responsabilità penale in questi contesti accentua le problematiche connesse alla rilevanza che il principio di precauzione ha nello stabilire la regola cautelare disattesa dal comportamento del soggetto ritenuto responsabile. Per contenere il diritto alla sperimentazione, espressione della più generale libertà d'azione, entro limiti che non nuocciano a terzi, senza ricorrere a un'eccessiva anticipazione della tutela penale, la soluzione potrebbe forse essere creare delle norme penali ad hoc per la fase di sperimentazione, che tengano conto della

---

<sup>72</sup> Si tratta di aspetti già approfonditi nel cap. I, cui si rinvia per le ulteriori precisazioni.

<sup>73</sup> F. Giunta, *La normatività, cit.*, p. 103.

situazione di forte incertezza in cui il fatto lesivo si è venuto a verificare<sup>74</sup>. Questo “protocollo di sperimentazione” renderà l’attività sperimentale lecita finché assistita da regole cautelari anch’esse sperimentali già verificate con successo per attività analoghe a quella intrapresa ovvero, nel caso in cui sia impossibile un raffronto con norme già collaudate, che risultino ex ante “altamente idone[ee] a contenere la pericolosità della nuova attività”<sup>75</sup>.

## 5. Rischio consentito

Nell’ambito di attività rischiose come quella dell’utilizzo e, prima ancora, della costruzione di sistemi intelligenti, la forza espansiva del diritto penale non sarà frenata solo dallo sbarramento che incontra negli ordinamenti moderni il principio di precauzione come tecnica di normazione. Un altro limite è rappresentato da quel compromesso tra esigenza di sviluppo tecnologico e esigenza di tutela dei beni giuridici. Queste due necessità sono in parte contrapposte, per altra parte invece si alimentano a vicenda: la tutela dei beni giuridici penalmente rilevanti potrà anche avvenire attraverso le tecnologie intelligenti, come già accade con l’impiego dell’intelligenza artificiale nelle attività di prevenzione del crimine da parte delle forze dell’ordine o nel settore giudiziario per la valutazione della pericolosità sociale di un individuo o del rischio di recidiva del reo; un utilizzo di queste tecnologie così orientato potrà far apprezzare i benefici dei sistemi di IA e favorire la ricerca nel settore.

Questa doppia faccia dell’intelligenza artificiale, che ne mostra la capacità potenziale di lesione o messa in pericolo di beni giuridici, per un verso, e la capacità di tutela di questi stessi beni e anche di incentivo per il soddisfacimento di altri interessi che esulano dalla materia penale, per altro verso, la accomuna a tutte quelle attività pericolose ma socialmente utili che l’ordinamento decide di consentire. Per risolvere la contraddizione in cui si sarebbe trovato l’ordinamento che da un lato autorizza queste attività, dall’altro

---

<sup>74</sup> È la soluzione sposata dal legislatore francese, che ha introdotto delle norme sulla responsabilità penale nel caso di reati verificatisi durante la sperimentazione di auto a guida autonoma con legge n. 2019-486 del 22 maggio 2019 (*Loi PACTE*). La legge ha autorizzato l’utilizzo sperimentale di auto a guida autonoma nelle vie pubbliche, previo rilascio di un’autorizzazione a fini sperimentali, e ha modificato l’art. 2-2 ord. n. 2016-1057 del 3 agosto 2016, prevedendo che la responsabilità del conducente è esclusa se il sistema è attivato ed è impiegato in modo conforme alle istruzioni per l’utilizzo; in presenza di tali circostanze, la responsabilità ricade allora sul titolare dell’autorizzazione di sperimentazione, sia per le violazioni delle norme del codice della strada, che per omicidio e lesioni stradali, ferme restando le condizioni di cui all’art. 121-3 c. pén. Per gli approfondimenti, si rinvia alla trattazione dedicata all’argomento nel cap. 4.

<sup>75</sup> Così, F. Giunta, *La normatività, cit.*, p. 104.

condanna i rischi che dall'esercizio delle stesse derivano, la dottrina tedesca ha elaborato la teoria del rischio consentito<sup>76</sup>.

La nozione di rischio penale è particolarmente vasta, poiché diversi sono i punti di emersione del rischio nella teoria del reato; il nucleo che accomuna le molteplici applicazioni del criterio del rischio è il porre la sanzione solo a partire da un certo livello di rischio, non essendo ammesso un intervento penale talmente esteso da giungere a vietare condotte che l'ordinamento non ha invece interesse a proibire<sup>77</sup>. Il rischio è quindi da intendersi come la *relazione di possibilità fra due entità penalmente rilevanti*, cioè la condotta umana e l'offesa –quest'ultima nella sua duplice veste di lesione o messa in pericolo di beni giuridicamente rilevanti-, che non è obiettivamente o naturalisticamente esistente, ma rileva già da un punto di vista *normativo*<sup>78</sup>.

L'aggettivo “consentito” è accostato al rischio per descrivere un'attività che entro certi limiti è valutata come penalmente irrilevante *dall'ordinamento nel suo complesso*<sup>79</sup> in ragione della sua elevata utilità sociale.

Spostando il fuoco del discorso verso le attività produttive, e richiamando la distinzione proposta in dottrina tra responsabilità per il tipo di produzione e per il modo di produzione<sup>80</sup>, è di immediata evidenza come la prima forma di responsabilità si leghi alle problematiche dell'individuazione dell'area del rischio consentito<sup>81</sup>. Si è detto, infatti, che la responsabilità per il tipo di produzione postula a monte un'attività produttiva

---

<sup>76</sup> Tra le opere fondamentali, Preuß, *Untersuchungen zum erlaubter Risiko im Strafrecht*, Berlin, 1974; H. Roeder, *Die Einhaltung des sozialadäquaten Risikos*, Berlin, 1969; Rehberg, *Zur Lehre vom „erlaubten Risiko“*, Zürich, 1962. Nella dottrina italiana, V. Militello, *Rischio e responsabilità penale*, Milano, Giuffrè, 1988; G. Forti, *Colpa ed evento*, cit., pp. 250 ss.; M. Gallo, voce *Colpa penale*, cit., p. 638; G. Marinucci, *La colpa*, cit., p. 210 ss.; M. Donini, *Lettura sistematica delle teorie dell'imputazione oggettiva dell'evento*, P.te I, in *Riv. it. dir. proc. pen.*, 1989, p. 588; F. Consulich, voce *Rischio consentito*, *Enc. dir., I Tematici, Il reato colposo*, Milano, Giuffrè, 2021, pp. 1102 ss.

<sup>77</sup> V. Militello, *Rischio e responsabilità penale*, Milano, Giuffrè, 1988, pp. 2 ss. L'A. studia le declinazioni che il rischio ha nella teoria del reato.

<sup>78</sup> *Ibidem*, pp. 33-34.

<sup>79</sup> Anche se uno dei campi in cui si è discusso originariamente dell'*erlaubtes Risiko* è anche quello dell'attività facoltizzata non dall'ordinamento, ma dal consenso della vittima che diviene situazione giustificante escludente la contrarietà alla norma penale. V. Militello, cit., pp. 37 ss. Una precisazione linguistica è doverosa. L'A. chiarisce che l'aggettivo “consentito” rispecchia la nozione che di tale istituto si è diffusa nella dottrina italiana, che lo ricollega alle attività tipicamente pericolose, ma autorizzate per la loro utilità sociale; nondimeno, una visione a più ampio raggio consente di appurare come l'istituto del rischio involge il reato anche in contesti diversi dalle attività produttive o pericolose, ricollegandosi alla necessità che ogni condotta umana abbia oltrepassato una soglia di pericolosità per essere penalmente rilevante; ebbene, per questo motivo l'A. preferisce impiegare la locuzione “rischio adeguato o lecito”, per riferirsi alla nozione ampia di rischio in materia penale e non ristretta al solo ambito delle attività connotate da una utilità sociale benché produttrici di offese (v. pp. 45 e 46).

<sup>80</sup> F. Bricola, cit. *supra*.

<sup>81</sup> Così, C. Piergallini, *Danno da prodotto*, cit., p. 41.

connotata da elevati livelli di pericolosità non neutralizzabile con il ricorso a misure cautelari ed è in questo contesto che l'ordinamento, valutando positivamente l'utilità della produzione, traccia l'area del rischio consentito.

La responsabilità del produttore di un sistema di IA ad alto rischio si collocherà allora nell'alveo della responsabilità per il tipo di produzione, trattandosi di prodotti che presentano un rischio intrinseco legato alle modalità di funzionamento (sistemi che procedono per generalizzazioni, in cui un margine di errore è sempre presente) e, per alcuni di essi, anche alle capacità di autoapprendimento (sistemi di *machine learning*). Di queste caratteristiche del bene è cosciente il produttore, ed è per questo che la dottrina ha messo in evidenza come, nel caso di autoapprendimento, si sia al cospetto di macchine ontologicamente imprevedibili, la cui connaturata imprevedibilità è però prevedibile dal costruttore<sup>82</sup>. Mi pare però che questa "prevedibile imprevedibilità" connoti tutti i sistemi di intelligenza artificiale che formulano una decisione individuando correlazioni statistiche o *pattern* nei dati di input. Qui la generalizzazione dalla quale è partorito l'*output* rende prevedibile un possibile errore del sistema basato su un modello statistico<sup>83</sup> (che può generare ad esempio falsi positivi o falsi negativi laddove il *task* affidato sia di classificazione), mentre l'imprevedibilità afferirà al *quando*, perché il costruttore non sarà in grado di conoscere *a priori* il momento e l'occasione in cui tali errori si verificheranno.

Si pone allora la necessità di interrogarsi sul grado di rischio che si è disposti a tollerare, definendo quali siano i rischi accettabili in base al tipo, all'impiego e al beneficio del sistema di IA<sup>84</sup>, sapendo che si tratta di una scelta politica<sup>85</sup> che non può essere rimessa al produttore. Ora è evidente che tale valutazione nel settore dell'intelligenza artificiale è espletata dall'ordinamento europeo, prima ancora che da quello nazionale.

L'Unione europea è cosciente di non poter rinunciare alla ricerca e alla sperimentazione in questo settore, così come alla produzione di beni intelligenti, e cerca allora di occuparsi di nuove fenomenologie di esposizione a pericolo di beni e valori rilevanti, posto che l'assoluta e incondizionata tutela dei beni giuridici non è praticabile in nessun ordinamento, se non a costo di paralizzare la vita sociale<sup>86</sup> che, in quanto vita

---

<sup>82</sup> C. Piergallini, *Intelligenza artificiale, cit.*, p. 1759.

<sup>83</sup> Come si è cercato di spiegare nel cap. 1, par. 2.2.2. *Bias e varianza*, cui si rinvia.

<sup>84</sup> Così, G. Ubertis, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Sistema penale 2020*, p. 9.

<sup>85</sup> C. Piergallini, *Intelligenza artificiale, cit.*, p. 1750.

<sup>86</sup> G. Forti, *Colpa ed evento, cit.*, p. 456.

di relazione, espone a rischio i beni che l'ordinamento tutela. È, quella europea, una legislazione consapevole di disciplinare un settore che vede nel rischio un elemento intrinseco, come dimostra la scelta di classificare i sistemi di IA ad alto e basso rischio, senza prendere in considerazione l'eventualità di distinguere tra sistemi di IA rischiosi e non. Piuttosto, i poli opposti sono tracciati all'interno della categoria di IA ad alto rischio, ove si distingue tra attività consentite e vietate<sup>87</sup>, facendo rientrare nel divieto sistemi il cui uso "è considerato inaccettabile, in quanto contrario ai valori dell'Unione"<sup>88</sup>. Si tratta di una scelta coerente con l'idea per cui nella definizione dell'area di rischio consentito giocano un ruolo importante il rango e la quantità dei beni giuridici esposti a pericolo<sup>89</sup>, il cui logico precipitato è che in settori in cui le conoscenze scientifiche non consentono ancora un controllo del pericolo adeguato al valore del bene esposto al rischio, ci si dovrà astenere dall'attività<sup>90</sup>: ecco così individuata l'area delle attività produttive ad alto rischio vietate dalla normativa europea.

La produzione di sistemi di intelligenza artificiale rappresenta l'ultima tappa, almeno per ora, della linea evolutiva del rischio consentito nel diritto penale, che vede ancora un percorso in crescente ascesa, andando di pari passo col progresso, che per sua natura tende verso l'incremento. Il terreno di studio del rischio lecito è, in questo contesto, l'illecito colposo. Con l'affermarsi della concezione normativa della colpevolezza, il rischio adeguato è confluito nella struttura del reato colposo soprattutto tramite l'elemento della diligenza<sup>91</sup>: il rischio consentito si inserisce nella fase genetica della regola cautelare imponendo di rileggere il criterio della riconoscibilità o rappresentabilità nel senso che dà luogo a colpa soltanto la condotta che si spinga oltre i limiti dell'adeguatezza sociale<sup>92</sup>. L'accertamento dell'ambito di rischio consentito e determinazione della figura modello sono, allora, aspetti collegati ma distinti, atteso che è possibile individuare una figura

---

<sup>87</sup> Cfr. proposta di regolamento dell'aprile 2021, in cui si individuano pratiche di IA vietate (Tit. II), sistemi ad alto rischio (Tit. III) e non ad alto rischio (Tit. IX).

<sup>88</sup> Par. 5.2.2. della proposta, p. 14 nella traduzione in lingua italiana. L'elenco delle attività vietate è riportato all'art. 5 della proposta. Tra queste, sono vietate le pratiche che utilizzano tecniche subliminali che agiscono senza che un soggetto sia consapevole, distorcendone il comportamento in modo da danneggiare, anche solo potenzialmente, soggetti terzi; o ancora sistemi che sfruttano le vulnerabilità di uno specifico gruppo di persone, sistemi a messi a disposizione della autorità pubblica finalizzate alla profilazione delle persone fisiche attraverso la pratica del c.d. *scoring*, sistemi di identificazione biometrica remota al di fuori dei casi espressamente autorizzati.

<sup>89</sup> Così, G. Forti, *cit.*, p. 308 e p. 459.

<sup>90</sup> *Ibidem*, p. 465.

<sup>91</sup> V. Militello, *cit.*, p. 74.

<sup>92</sup> M. Gallo, *cit.*, p. 640.

modello che sia espressione del rischio consentito, in quanto tale ambito esista e sia individuabile<sup>93</sup>.

Il legame tra rischio consentito e diligenza doverosa è espresso dalla distinzione, proposta dalla dottrina, tra regole cautelari proprie e improprie<sup>94</sup> alla luce dell'efficacia impeditiva del rischio per effetto del comportamento conforme alla regola. Alla base della distinzione risiede l'idea per cui la componente rappresentativa della regola cautelare (*i.e.* la riconoscibilità dell'evento) si caratterizza per “una sostanziale “fissità””, nel senso che “l'evento pregiudizievole o è pregiudizievole, o non lo è”<sup>95</sup>; ciò non vale invece per la componente dell'evitabilità dell'esito lesivo, che presenta un grado variabile che determina la distinzione tra regole cautelari proprie e improprie.

Sono *proprie* le regole cautelari che prescrivono una condotta in grado di evitare con *una probabilità confinante con la certezza* l'evento. Tale grado di elevata probabilità manca, invece, nel caso delle regole cautelari improprie, che impongono un contegno che è in grado di prevenire l'evento, ma senza che sia possibile nutrire alcuna ragionevole certezza in merito<sup>96</sup>. Naturalmente, la regola cautelare impropria è accettabile se calata in un contesto di attività socialmente utili, per questo autorizzate o consentite pur in presenza di un margine di rischio ineliminabile<sup>97</sup>, ma consentito.

Ben si comprendono le perplessità legate all'accoglimento di un ruolo delimitativo della diligenza delle regole cautelari improprie affidato al rischio consentito: la genericità di siffatto criterio non soddisfa l'esigenza di determinatezza di cui si riveste l'illecito penale, in ossequio al principio di legalità, ma anche di colpevolezza, atteso che nessun rimprovero servirebbe a rieducare chi non riconosce il disvalore della propria condotta perché non è messo nelle condizioni di farlo<sup>98</sup>.

Soddisfare l'esigenza di determinatezza del criterio del rischio adeguato lasciando più spazio nell'individuazione dello stesso alle autorizzazioni amministrative è la proposta che è emersa in dottrina, ma non si è mancato di evidenziarne i limiti, atteso che l'ancoraggio alle autorizzazioni amministrative pone sul tavolo le problematiche tipiche della colpa specifica, rientrando le prime tra le “discipline” di cui si fa menzione nell'art.

---

<sup>93</sup> Così, G. Forti, *cit.*, p. 307.

<sup>94</sup> P. Veneziani, *Regole cautelari “proprie” ed “improprie”*, Padova, Cedam, 2003.

<sup>95</sup> Così, P. Veneziani, *cit.*, p. 16.

<sup>96</sup> Così, P. Veneziani, *cit.*, p. 15.

<sup>97</sup> *Ibidem*, p.16.

<sup>98</sup> La critica dei mobili confini del concetto del rischio consentito è da riferire, nella dottrina italiana, a M. Gallo, *cit.*; G. Marinucci, *La colpa, cit.*, p. 210; G. Fiandaca, E. Musco, *Diritto penale, cit.*, p. 587.

43 c.p.<sup>99</sup>. Così, come ogni regola cautelare scritta, anche per le autorizzazioni va segnalata l'incapacità di estromettere le fonti sociali dal compito di conformare il comportamento dell'agente: la colpa specifica è ricavata all'interno del più ampio insieme della colpa generica, cosicché quest'ultima tornerà a operare oltre i confini della regola di diligenza scritta come "colpa generica residuale"<sup>100</sup>. Appiattare l'accertamento della colpa alla sola conformità della produzione ai presupposti autorizzativi potrebbe disincentivare l'imprenditore a produrre beni sempre più sicuri e ad allinearsi alle scoperte scientifiche più recenti<sup>101</sup>. Per questo, il produttore, al quale ragionevolmente saranno indirizzate le autorizzazioni amministrative, non potrà giustificarsi dimostrando la perfetta aderenza del proprio contegno alla regola posta dall'Autorità se poi ha disatteso un obbligo di fonte sociale<sup>102</sup>.

I limiti qui presentati spingono allora a sostenere che il bilanciamento tra i diversi interessi che conosce la sua sintesi nel criterio del rischio consentito non può essere rimesso sempre all'Autorità pubblica conferendole un potere autorizzatorio, anche perché, a portare alle estreme conseguenze questo pensiero, si finirebbe per vietare tutto ciò che non è espressamente autorizzato, con un evidente limite per le libertà del cittadino.

Deve allora ritenersi che, pur riconoscendo l'importante ruolo delle autorizzazioni, l'individuazione del rischio consentito è un processo cui partecipa l'intero ordinamento<sup>103</sup>, attraverso l'opera congiunta del legislatore e dell'interprete. Il ruolo del giudice, però, non è mai libero da vincoli legali. Anche là dove manchino norme puntuali anche di rango secondario che disciplinino l'attività coinvolta nell'illecito colposo, il giudice sarà chiamato a interpretare la clausola del rischio adeguato alla luce dell'ordinamento nel suo complesso considerato, partendo già dalla Carta fondamentale, che offre una sintesi del bilanciamento tra sviluppo economico e tutela della persona umana nell'art. 41 Cost. Leggendo il rischio consentito alla luce di tale disposizione, esso non apparirà mai una "scatola vuota", ma si presenterà piuttosto come canone inserito nel tessuto

---

<sup>99</sup> V. Militello, *cit.*, pp. 94 ss.

<sup>100</sup> Così, F. Giunta, *La normatività, cit.*, p. 93.

<sup>101</sup> Così, V. Militello, *cit.*, p. 96.

<sup>102</sup> Il rapporto tra colpa specifica e generica è spiegato, tra gli altri, da F. Mantovani, *voce Colpa, cit.*, p. 309; F. Giunta, *La normatività, cit.*, p. 93-94, il quale sottolinea che la colpa generica residuale potrà sussistere per rischi diversi da quelli fronteggiati dalla regola cautelare positivizzata. A ritenere diversamente, quindi ad ammettere una colpa generica per quei profili di pericolosità della condotta riconducibili allo spettro preventivo della cautela scritta disattesa, si avallerebbe la creazione in via analogica di nuove fattispecie colpose, attribuendo all'interprete un potere creativo che non gli è riconosciuto dall'ordinamento.

<sup>103</sup> Ancora, V. Militello, *cit.*, p. 112.

dell'ordinamento vigente, che però per operare avrà sempre bisogno di un intervento giudiziale<sup>104</sup>. In questo modo la clausola del rischio consentito diventa espressione del contemperamento tra generale principio di libertà dell'agire umano e la tutela dei beni giuridici<sup>105</sup>. Si tratta di una soluzione non nuova nel panorama della colpa: di indeterminatezza è talvolta vestito già il parametro normativo della negligenza, che spetta al giudice concretizzare<sup>106</sup>, ricorrendo a una discrezionalità che non è mai libera e che, quindi, non è mai arbitrio.

Nel settore della IA resta quindi prioritaria, in un ordinamento ispirato al principio di legalità, la necessità di tracciare a livello legislativo i margini tra lecito e illecito, ricorrendo anche a tecniche di normazione integrata da fonti secondarie più facilmente aggiornabili e in grado di seguire il progresso nella materia dell'IA. Il ruolo del legislatore non sottrae però spazio all'attività dell'interprete, il quale ha un rapporto stretto con il fatto concreto ed è l'unico cui può essere rimesso il bilanciamento finale tra le esigenze che la nozione di rischio consentito riassume. Il giudizio si muoverà comunque nei binari tracciati dal legislatore, al quale è richiesto di tracciare il percorso regolamentativo in un settore di produzione che porta con sé nuovi rischi.

Finora tale compito di regolazione è stato assunto a livello europeo, dove il divieto di talune pratiche di intelligenza artificiale, nello specifico quelle indicate all'art. 5 della proposta di regolamento, rappresenta chiaramente una prima delimitazione del confine tra il lecito e l'illecito.

All'interno del gruppo di sistemi ad alto rischio è poi prevista una procedura di valutazione di conformità dei prodotti da parte di appositi organismi di valutazione, che acquisiscono tale potere valutativo solo dopo essere stati a ciò autorizzati dall'Autorità di notifica, ente che sarà istituito in ciascuno Stato membro e ai quali le aziende che hanno capacità di valutare la sicurezza dei prodotti di IA possono rivolgere la domanda di notifica (artt. 30-32).

Il legislatore europeo instaura una gerarchia di controlli, affidata a privati che abbiano ottenuto la fiducia dell'autorità pubblica, dimostrando di essere competenti nel valutare la conformità di determinate tecnologie di IA (art. 31).

---

<sup>104</sup> *Ibidem*, p. 116 ss.

<sup>105</sup> V. Militello, *cit.*, p. 123.

<sup>106</sup> *Ibidem*, p. 115.

Già dall'impostazione data a livello europeo, non v'è dubbio che la legislazione degli Stati membri si impegnerà a precisare ulteriormente i limiti entro i quali il rischio sarà consentito e, forse, dovrà essere evitato il problema opposto, cioè quello di una eccessiva burocratizzazione della produzione di sistemi intelligenti, che possa anche prestare il fianco a dinamiche corruttive nei rapporti con gli organismi deputati all'attività di valutazione e con la stessa Autorità di notifica, le quali svolgerebbero poteri certificativi che ben potrebbero rientrare nella nozione di pubblico servizio di cui all'art. 358 c.p.

## **6. La posizione del produttore. La perimetrazione della colpa specifica**

Nel contesto di interesse per questo studio, la regola cautelare si rivolgerà al produttore e all'utente e potrà eventualmente essere positivizzata così ponendosi le basi per il rilievo della colpa specifica.

Come noto, la fonte delle regole alla base della colpa specifica<sup>107</sup> può essere legislativa o regolamentare ovvero avere la forma di un ordine o di una disciplina; l'atto di normazione può promanare da un'autorità pubblica, ma anche privata nel caso di discipline e ordini, avere carattere generale o, solo nel caso degli ordini, individuale (si pensi alle norme poste dal datore di lavoro pubblico o privato nell'ambito del rapporto di lavoro)<sup>108</sup>.

La tendenza a positivizzare le regole di condotta è sempre più marcata nell'era tecnologica, mentre le regole non scritte sono proprie di "epoche tecnologicamente più tranquille e riposanti"<sup>109</sup>. A tale tendenza non si sottrae il contesto produttivo delle tecnologie di IA, che conosce già una normazione europea che pone dei criteri da seguire nella catena produttiva, cosicché emerge già la propensione verso la positivizzazione delle regole destinate al produttore.

Nel semenzaio normativo eurounionale, un ruolo peculiare è quello delle dichiarazioni contenenti i principi condivisi a livello europeo che, sebbene finalizzati a orientare la produzione dei sistemi di IA (si pensi ai principi di controllo umano, robustezza, trasparenza, esplicabilità, non discriminazione che ricorrono come *leit motiv*

---

<sup>107</sup> Cfr. R. Bartoli, voce *Fonti della colpa*, *Enc. dir., I Tematici, Il reato colposo*, Milano, Giuffrè, pp. 519 ss.

<sup>108</sup> S. Canestrari, *La colpa*, cit., p. 144; F. Mantovani, voce *Colpa*, cit., p. 306.

<sup>109</sup> F. Mantovani, voce *Colpa*, cit., p. 306.

in più atti delle istituzioni europee), proprio per il loro carattere generico, non possono rivolgersi a un fruitore finale di norme come il produttore. Tali dichiarazioni rappresentano piuttosto uno standard di normazione per il legislatore europeo, sicché i principi in esse contenuti non possono essere annoverati tra le regole codificate in grado di qualificare la colpa come specifica.

Una regolamentazione più precisa e puntuale è invece quella della proposta di regolamento dell'aprile 2021, che contiene regole che disciplinano con finalità cautelare la produzione dei sistemi di IA ad alto rischio, destinate quindi ai produttori. Esse sono contenute nel Capo II e III del Titolo I. Nondimeno, trattandosi solo una proposta di regolamento, allo stato non può ancora dirsi esservi regole cautelari positivizzate, poiché non ancora vigenti.

Il Capo II, nell'indicare i "requisiti giuridici dei sistemi ad alto rischio", individua già degli obblighi piuttosto puntuali rivolti al produttore. Ad esempio, si prevede che i prodotti intelligenti ad alto rischio siano dotati di sistemi di gestione dei rischi per l'intero ciclo di vita del sistema (art. 9), che consenta non solo di individuare i rischi *noti e prevedibili*, ma anche di adottare *adeguate misure di gestione* degli stessi (par. 2 lett. d)). Il produttore è quindi tenuto a eliminare o ridurre i rischi attraverso un'adeguata progettazione e fabbricazione (par. 4, lett. a)) e a predisporre misure di attenuazione e controllo di quei rischi che non possono essere eliminati (par. 4, lett. b)). Se qui non si erra nell'interpretare il significato di tale disposizione, sembra che il legislatore europeo imponga al produttore di introdurre nel sistema forme di controllo di emergenza dei rischi controllabili e forme di attenuazione del rischio incontrollabile. In questo modo si danno indicazioni sulle caratteristiche del prodotto da mettere in commercio, che deve sempre essere accompagnato da una sorta di "*first aid kit*" fornito dal produttore all'utente, utile nel caso in cui dei rischi si manifestino durante l'utilizzo del prodotto intelligente<sup>110</sup>.

Dei criteri di produzione si introducono anche con riguardo ai set di dati di addestramento, che devono soddisfare degli standard di qualità determinati (art. 10). Ancora, è richiesta la redazione di una dettagliata documentazione tecnica (art. 11), e che la progettazione garantisca un funzionamento trasparente dei sistemi, i cui output devono essere interpretabili dall'utente (art. 13). I sistemi ad alto rischio devono anche essere

---

<sup>110</sup> Vi sarà una "colpa di programmazione" in capo allo sviluppatore o programmatore per non aver previsto un intervento manuale sul sistema in condizioni di emergenza, così, I. Salvadori, *cit.*, p. 103.

robusti, cioè resistenti agli attacchi da parte di terzi volti a modificarne l'uso o le prestazioni sfruttando le vulnerabilità del sistema (art. 15, par. 4).

Regole sulla produzione sono contenute anche nel capo III. Si pensi, ad esempio, alla previsione di una procedura di valutazione di conformità del bene prima dell'immissione sul mercato (art. 19).

Come è dato notare, da queste disposizioni sulle caratteristiche dei sistemi possono già essere estrapolati dei doveri di diligenza nella produzione dei sistemi ad alto rischio, volti a ridurre i rischi che possono emergere dall'uso dei sistemi di IA. A fronte di questo apparato normativo, per ora contenuto però solo in una proposta di regolamento, possiamo quindi interrogarci sulla sua capacità di introdurre regole cautelari atte a qualificare la colpa come specifica piuttosto che generica.

A tal proposito va notato che non ogni regola scritta è per ciò solo in grado di fondare un rimprovero a titolo di colpa specifica, essendo piuttosto necessario che la norma positivizzata presenti certe caratteristiche, che guidano una sorta di "procedimento per esclusione" all'esito del quale l'ambito della colpa specifica risulta fortemente ridimensionato<sup>111</sup>.

La restrizione del raggio di azione della colpa specifica è un logico precipitato delle caratteristiche di questa tipologia di colpa, che esprime e sintetizza giudizi di prevedibilità (o riconoscibilità) e evitabilità effettuati in astratto dal legislatore, cosicché l'interprete dovrà limitarsi ad accertare che il contegno dell'agente abbia realizzato il comportamento vietato ovvero non abbia tenuto quello richiesto. La rappresentabilità e l'evitabilità sono allora già "assorbite" nella regola giuridica e si presentano sotto forma di presunzione che l'interprete non può né sindacare in astratto (giudicando tale presunzione scorretta), né modellare alla luce delle circostanze del caso concreto<sup>112</sup>, dovendo il giudice piuttosto limitarsi ad accertare che la condotta contrasti con la regola prudenziale positivizzata. Il rischio, evidente, è quello di dare spazio a giudizi ispirati alla logica del *versari in re illicita*, nei quali l'evento è addossato all'agente a titolo di colpa per il solo fatto di avere agito in contrasto con la regola precauzionale<sup>113</sup>. Questa "protocollizzazione" della colpa<sup>114</sup> è resa ragionevole dai benefici che, almeno in astratto, derivano da una disciplina

---

<sup>111</sup> O. Di Giovine, *Il contributo della vittima nel delitto colposo*, Torino, Giappichelli, 2003, p. 399.

<sup>112</sup> G. Marinucci, *La colpa per inosservanza*, cit., pp. 228-229.

<sup>113</sup> F. Mantovani, *voce Colpa*, cit., p. 308.

<sup>114</sup> O. Di Giovine, *Il contributo*, cit., p. 388.

scritta in termini di chiarezza e certezza<sup>115</sup>, cosicché la positivizzazione della regola di diligenza è particolarmente indicata allorché si voglia sottrarre la pretesa di diligenza alle variabili che caratterizzano l'esperienza individuale quando si tratti di sperimentare nuove tecniche sulle quali lo standard di esperienza maturato è ancora insufficiente<sup>116</sup>.

Ebbene, proprio i risvolti processuali e sostanziali della colpa specifica giustificano gli sforzi della dottrina nel delimitarne l'ambito di operatività, per affermare che non ogni regola scritta è in quanto tale idonea a delineare una colpa specifica, occorrendo elementi ulteriori che ne restringono la portata<sup>117</sup>.

Iniziamo allora innanzitutto notando che quel “pieno ‘assorbimento’ nella struttura della norma del calcolo di prevedibilità compiuto in anticipo dalla fonte” che ha posto la regola cautelare si ha solo nel caso di norme a struttura rigida, così definite perché non si prestano ad alcuna apertura sulla realtà concreta<sup>118</sup>.

Queste regole vengono contrapposte alle regole elastiche, così definite in quanto la fonte si limita solo a indicare lo scopo cautelare, restando invece silente in ordine alle condotte da tenere per la sua realizzazione, cosicché la positivizzazione della regola non assicura quell'elevato standard di precisione<sup>119</sup> che in astratto la colpa specifica promette. Il vincolo giuridico posto dalla regola elastica riguarda il raggiungimento dello scopo preventivo, restando sostanzialmente libere le concrete modalità cautelari di adempimento di tali aspettative, cosicché la pretesa cautelare acquista un contenuto dinamico<sup>120</sup>: con un linguaggio mutuato dal diritto civile, diremmo che sono fonti di obbligazioni di risultato e non di mezzi. In presenza di regole elastiche, allora, è impossibile prescindere dal giudizio di riconoscibilità e evitabilità in concreto alla luce del parametro dell'*homo eiusdem professionis et condicionis*, per cui bisogna immancabilmente appurare se le circostanze del caso concreto lasciavano prevedere la realizzazione del risultato dannoso<sup>121</sup>.

---

<sup>115</sup> G. Marinucci, *La colpa*, cit., p. 233. La dottrina non manca però di sottolineare come l'allineamento della colpa col principio della riserva di legge per effetto della eterointegrazione della fattispecie non è fenomeno esclusivo della colpa specifica, in quanto “anche nella colpa generica abbiamo regole cautelari che sia l'agente, sia il giudice trovano e non creano. Si tratta degli usi cautelari di natura consuetudinaria”, così F. Giunta, *Culpa, culpa*, in *disCrimen*, 4.06.2019, p. 9.

<sup>116</sup> Così, F. Giunta, *La normatività*, cit., p. 92.

<sup>117</sup> O. Di Giovine, *Il contributo*, cit., p. 389.

<sup>118</sup> G. Marinucci, *La colpa*, cit., p. 243.

<sup>119</sup> F. Giunta, *Culpa, culpa*, cit., p. 10.

<sup>120</sup> Così, D. Castronuovo, *L'evoluzione teorica*, cit., p. 1618.

<sup>121</sup> G. Marinucci, *La colpa*, cit., p. 239.

Senonché, la dottrina osserva che anche nell'intelaiatura delle norme rigide si notano aperture che escludono il riconoscimento della colpa specifica allorché la loro funzione preventiva sia neutralizzata dal caso concreto<sup>122</sup>. Ciò si verifica quando il rispetto del comando o del divieto della regola rigida comporterebbe una minaccia per il bene giuridico, cosicché è la stessa funzione di prevenzione della regola cautelare a imporre all'agente concreto di tenere una condotta diversa da quella nella stessa descritta. Ne risulta che anche le norme rigide vanno lette come recanti una clausola di riserva implicita per la quale la loro osservanza è richiesta fintantoché da essa non derivi un rischio per i beni tutelati. Per effetto di tale riserva, allora, l'obbligo generale di diligenza e prudenza che caratterizza la colpa generica subentra anche negli schemi della colpa specifica<sup>123</sup>.

Ma vi sono altre ipotesi individuate dalla dottrina<sup>124</sup> in cui la presenza della regola scritta richiede comunque l'accertamento della colpa generica. È questo il caso delle regole prudenziali definite di secondo grado, tendenti a prevenire la messa in pericolo di un bene e non un determinato evento specifico che lo coinvolga, cosicché l'eventuale offesa darà luogo a responsabilità colposa solo all'esito di una valutazione concreta della riconoscibilità e evitabilità dell'evento<sup>125</sup>. Le stesse esigenze in termini di accertamento della colpa generica derivano dalle regole con finalità amministrativa che perseguono però anche una finalità cautelare, come quelle che impongono la revisione dei veicoli alla luce dei criteri amministrativi imposti, col fine ultimo di tutelare l'incolumità del conducente e dei terzi, ovvero la norma che impone il conseguimento della patente di guida. Ancora, problematiche simili si porranno per quelle regole di condotta che si prefiggono di monitorare un certo rischio, non essendovi conoscenze scientifiche certe che permettano di individuare i comportamenti che possano piuttosto ridurlo o eliminarlo, per le quali l'addebito a titolo di colpa non può prescindere dall'accertamento in concreto della riconoscibilità e evitabilità dell'evento. In altri casi, la scienza può fornire risposte

---

<sup>122</sup> G. Marinucci, *La colpa*, cit., pp. 247-248.

<sup>123</sup> Osserva G. Marinucci (*Ibidem*) che tale riserva talvolta viene espressamente formulata dalla legge, come avviene nel caso dell'art. 140 cod. strada (precedentemente art. 101, come indicato dall'A. in riferimento al testo allora vigente), sul "*principio informatore della circolazione*", a mente del quale "*Gli utenti della strada devono comportarsi in modo da non costituire pericolo o intralcio per la circolazione ed in modo che sia in ogni caso salvaguardata la sicurezza stradale*". Si tratta, secondo l'A., di una norma di portata generale che funge da limite alla operatività di tutte le norme del codice, anche di quelle a struttura rigida.

<sup>124</sup> Il riferimento è all'analisi fatta da O. Di Giovine, *Il contributo*, cit., pp. 391 ss.

<sup>125</sup> Una regola cautelare di secondo grado è ad esempio quella che fa divieto di utilizzare il telefono alla guida, che mira a evitare tutta una serie di eventi indeterminati che possano scaturire dalla situazione di pericolo creata dal calo dell'attenzione provocato dall'uso del telefono, *Ibidem*, p. 392.

in ordine alla verificabilità o meno di un certo rischio che si traspongono in una certa regola cautelare, ma potrà darsi l'evenienza che quelle conclusioni siano poi superate da conoscenze sopravvenute, per cui al momento dell'accertamento della tipicità colposa la valenza della regola cautelare scritta sarà smentita dal progresso scientifico, che richiederà una valutazione in concreto, alla stregua del binomio riconoscibilità-evitabilità in concreto.

Da ultimo, anche la fonte della regola potrà giocare un ruolo nel richiedere un accertamento giudiziale che abbia riguardo al caso concreto e rinunci agli automatismi della colpa specifica, soprattutto laddove la fonte non abbia rango legislativo e non promani da un organo pubblico (è questo il caso degli ordini), bensì dallo stesso destinatario che interviene a autoregolamentarsi con una fonte inquadrabile nelle "discipline". L'autonormazione riduce l'efficacia preventiva della regola precauzionale, che non può essere considerata assoluta e per questo è sindacabile da parte del giudice<sup>126</sup>.

Orbene, questa classificazione delle regole scritte, che ritaglia i bordi della vera colpa specifica per individuare i casi in cui non si può prescindere dall'accertamento concreto della tipicità colposa tipico della colpa generica, è utile nell'analisi che stiamo conducendo in ordine alla responsabilità del produttore che, non è da escludere, in molte situazioni richiederà l'accertamento della colpa alla luce del parametro dell'*homo eiusdem professionis et condicionis* che caratterizza la colpa generica. E infatti, già dallo studio della proposta di regolamento emergono molte regole cautelari che potremmo classificare come di secondo grado, volte cioè a prevenire dei rischi e non degli eventi determinati (penso possa essere ricondotta entro questa categoria la maggior parte delle regole sopra enunciate, a partire dall'art. 9, sul sistema di gestione dei rischi, per passare poi a quella sui requisiti dei set di dati di cui all'art. 10, e concludendo con quelle dell'art. 13 sulla trasparenza e interpretabilità dell'output e dell'art. 15 sulla robustezza dei sistemi), oltretutto norme con finalità di monitoraggio e norme con attitudine amministrativa e finalità cautelare solo in via mediata (è questo il caso del regime delle certificazioni e della valutazione di conformità). Infine, è certo che ogni impresa si doterà di un proprio corpus normativo o che esso sia fornito da soggetti privati esterni all'azienda

---

<sup>126</sup> Ancora O. Di Giovine, *Il contributo*, cit., p. 399.

che svolgono il ruolo di consulenza<sup>127</sup>. Si tratterà allora di discipline che pongono regole cautelari.

Ebbene, in tutti questi casi il giudice non potrà limitarsi ad accertare che la condotta del produttore si discosta dal contenuto della regola cautelare scritta, col risultato che diventa imprescindibile un accertamento che rivolga la lente al caso concreto e alla riconoscibilità e evitabilità dell'evento che poteva avere un agente ideale immaginato al posto di quello concreto. Sarà cioè inevitabile il ricorso al parametro del produttore modello pur in presenza di indizi che ci faranno in prima battuta pensare che si tratti di una colpa specifica.

## 6.1 Produttore modello e prodotto modello

A questo punto risulta chiaro che la proliferazione di regole sulla produzione di sistemi di IA cui si assisterà nel prossimo breve periodo non produrrà in via automatica l'ampliamento della colpa specifica, il cui ambito sarà comunque ridimensionato per effetto del contenuto delle regole stesse. Inoltre, anche laddove una effettiva colpa specifica sia individuata all'esito di quel 'processo per esclusione' di cui si è dato atto, non va dimenticato che residueranno comunque margini per la valutazione della rilevanza penale della condotta del produttore sulla scorta di un giudizio sulla riconoscibilità e evitabilità in concreto. Questi rilievi indirizzano la ricerca a soffermarsi sulla figura del "produttore modello", che rappresenterà il parametro per valutare la risposta comportamentale dell'agente concreto al fine di vagliarne la corrispondenza alla tipicità colposa. Si può allora sostenere che il produttore diligente è colui che costruisce un prodotto che presenti le caratteristiche indicate dalla legge e che sia sicuro, cioè che non esponga a rischio beni presidiati dallo strumento penale. Ma riflettiamo: così inteso, il parametro del produttore modello sarà plasmato sul "prodotto modello" -cioè il prodotto

---

<sup>127</sup> Tra gli atti che si prestano a divenire future "discipline" ai fini del giudizio sulla colpa mi pare possa essere annoverato a titolo di esempio il report dal titolo "*capAI. A procedure for conducting conformity assessment of AI systems in line with the EU Artificial Intelligence Act*", pubblicato il 23 marzo 2022 (versione 1.0). Si tratta di una procedura di valutazione della conformità (da cui l'acronimo cap dall'inglese "conformity assessment procedure") della produzione dei sistemi di IA, a partire proprio dai principi fissati dalla proposta di regolamento, messo a punto da un gruppo di esperti del Centro per l'Etica Digitale (CEDE) attivo al Dipartimento di Scienze Giuridiche dell'Università di Bologna, della Saïd Business School e dell'Oxford Internet Institute dell'Università di Oxford. Floridi, Luciano and Holweg, Matthias and Taddeo, Mariarosaria and Amaya Silva, Javier and Mökander, Jakob and Wen, Yuni, *capAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act* (March 23, 2022). Available at SSRN: <https://ssrn.com/abstract=4064091> or <http://dx.doi.org/10.2139/ssrn.4064091>.

che rispetta le caratteristiche indicate dalla legge e che sia al contempo sicuro-, cosicché il raffronto sarà da intendersi tra prodotti e non tra produttori. Se è produttore modello colui che costruisce un bene senza difetti, o quantomeno, conforme alle caratteristiche richieste dalla legge, allora per accertare la colpa ci chiederemo se il bene prodotto in concreto rispetti gli standard del bene privo di difetti. Cambia, quindi, il parametro di riferimento: da agente modello a prodotto modello.

Certo, il parametro del “prodotto modello” non sostituisce quello del “produttore modello”, ma lo caratterizza per ciò che concerne il profilo della responsabilità per i danni da prodotto. Le caratteristiche del prodotto saranno invece irrilevanti se il produttore sarà ad esempio considerato nelle vesti di datore di lavoro nel contesto della responsabilità per i danni alla salute dei propri dipendenti. È quindi nel solo ambito del danno da prodotto che emerge la correlazione tra “produttore modello” e “prodotto modello”.

Ancora diversa è la prospettiva laddove si sostenga l’idea di una responsabilità singolare del sistema di IA: in tale evenienza il parametro del ‘prodotto modello’ si sostituirà a quello del produttore modello o i due rileveranno in modo concorrente per profilare due tipi di responsabilità, cioè quella classica del produttore persona fisica e quella singolare del sistema di IA inteso quale entità destinataria di regole e sanzioni, tra cui anche quella penale. Se infatti l’autore del fatto criminoso è considerato il sistema intelligente al quale viene riconosciuta una personalità giuridica, il suo agire sarà considerato negligente alla luce di un comportamento ideale che è quello del “prodotto modello”, cioè del bene che è privo di difetti e che quindi non cade in errore durante il processo decisionale<sup>128</sup>.

Come si è già detto in apertura di questo capitolo, questo itinerario di responsabilità rappresentato dalla responsabilità singolare del sistema di IA non è al momento considerato percorribile nel contesto europeo, né tanto meno potrebbe esserlo nel campo di questa ricerca, cioè quello della responsabilità penale, dove non poche sarebbero le difficoltà che si incontrerebbero nell’individuare nel sistema intelligente un nuovo centro di imputazione della responsabilità.

---

<sup>128</sup> Riflette su questo tema R. Abbot, *The Reasonable Robot. Artificial Intelligence and the Law*, Cambridge, Cambridge University Press, 2020, p. 69, seppur dalla prospettiva dell’autore del danno (e non da quella del produttore), (“*AI tortfeasors should be held to a negligence standard, and their acts ought to be compared to those of a reasonable person. Further, the reasonable person standard should be replaced with the reasonable robot standard once automation is practicable and AI is safer than the average person. Eventually, this means that AI tortfeasors will be held to the reasonable robot standard. For instance, if a self-driving Audi collides with a child who ran in front of the vehicle, the negligence test could take into account the stopping times of self-driving Google cars*”).

Torniamo allora a occuparci della responsabilità per i danni cagionati durante l'uso di un sistema intelligente seguendo i binari classici della responsabilità penale personale riferita a una persona fisica. Se si condivide quanto si è detto circa la correlazione tra la figura del produttore modello e il prodotto modello, allora significa che per i danni da prodotto la diligenza o negligenza del produttore vanno accertate considerando le caratteristiche del prodotto e eventuali difetti che hanno determinato un cattivo funzionamento fonte di danno.

Tracciata quindi la nostra linea di indagine per la responsabilità del produttore, bisogna soffermarsi sulle caratteristiche che il prodotto intelligente deve presentare per essere privo di difetti.

## 6.2 Diligenza artificiale

Si è sostenuto che il livello di diligenza del produttore è stabilito anche alla luce del “prodotto modello”, cioè del bene privo di difetti. Considerato che i sistemi intelligenti si caratterizzano per la capacità di assumere delle decisioni, potremmo ritenere che anche gli artefatti tecnologici tengono durante il loro utilizzo un “comportamento”, che consiste nella produzione di un output a partire da un certo input. Orbene, questo “comportamento” del sistema non è lasciato al caso, ma è guidato dalle informazioni inserite nello stesso durante la fase di progettazione.

Nel caso di prodotti ‘agenti’ cioè che si muovono nello spazio e assumono decisioni, come fanno i sistemi intelligenti, potremmo allora dire che il bene privo di difetti è quel manufatto che “agisce con diligenza”<sup>129</sup>. È chiaro che questa diligenza non è assimilabile

---

<sup>129</sup> L'idea di un parametro di diligenza riferito direttamente al sistema artificiale è presa in considerazione nell'ambito della responsabilità civile, dove si suole fare riferimento al c.d. “reasonable algorithm” con riguardo agli algoritmi di ML, che assumono autonomamente delle decisioni. Fermo restando che lo standard del ‘reasonable algorithm’ troverà sicuro spazio applicativo allorché sia riconosciuta agli algoritmi una personalità giuridica, venendo quindi equiparati all'uomo (v. R. Abbot alla nota precedente), si ritiene che esso possa avere rilevanza anche laddove l'algoritmo non sia considerato un autonomo centro di imputazione, poiché tale valutazione serve a indirizzare il giudizio sul comportamento del produttore verso una forma di responsabilità che richiede un onere probatorio diverso da quello necessario nella responsabilità da prodotto difettoso; la diligenza del produttore andrebbe piuttosto valutata alla luce dello standard del reasonable algorithm, K. Chagal-Feferkorn, *The Reasonable Algorithm*, in *University of Illinois Journal of Law, Technology & Policy*, 2018, no. 1 (Spring 2018), pp. 111-148. Ma la dottrina straniera non è concorde sul tema: si registrano infatti opinioni volte a evidenziare come non sia necessario vagliare la ragionevolezza (diligenza) algoritmica, posto che gli algoritmi non sono equiparabili agli agenti umani, ma restano comunque degli strumenti utilizzabili dagli umani, K. Colonna, *Autonomous Cars and Tort Liability*, 4 CASE W. RES. J. L. TECH & INTERNET 81, 102- 04 (2012); J. M. Balkin, *The Three Laws of Robotics in the Age of Big Data*, 78 *Ohio St. L.J.* (Sep. 10, 2017), che evidenzia che “there is no little person inside the program” e che gli algoritmi adottano le decisioni per le quali sono stati programmati,

a quella umana, perché la macchina, seppur intelligente, è diversa dall'uomo. Questa è una diligenza "artificiale", nel senso che deriva dalle conoscenze che il sistema di IA ha acquisito durante le fasi di produzione, ovvero dalle informazioni apprese durante il suo utilizzo.

Può darsi che tali informazioni coincidano con regole a contenuto precauzionale che, così come orientano il comportamento degli agenti umani, sono destinate a indirizzare il comportamento dell'agente artificiale chiamato a svolgere una funzione tipicamente destinata all'uomo. Si pensi alla condotta di guida su strada o a quella di diagnosi medica. Il guidatore, prima di mettersi alla guida, è chiamato a conoscere le regole del codice della strada; allo stesso modo, il medico prima di effettuare una diagnosi, è tenuto a conoscere le regole che consentono di addivenire a una corretta diagnosi e a comportarsi in sostanza nel modo previsto dalle *leges artis*.

Nel momento in cui tali compiti vengono demandati a un sistema di IA, una simile pretesa di conoscenza delle regole del codice della strada e delle *leges artis* è rivolta non più all'agente umano, ma a quello artificiale. Da ciò deriva che in tale contesto la regola precauzionale diviene elemento conoscitivo che va inserito nel sistema intelligente attraverso l'addestramento e che contribuisce a formare la conoscenza dello stesso.

### **6.2.1 Base ontologica e base nomologica della riconoscibilità artificiale**

Potremmo a questo punto definire *diligente* il comportamento del sistema che *riconosce* una data situazione attraverso la corretta interpretazione dei dati di input e che produce un output conforme a quello richiesto; per converso, sarà *negligente* il comportamento del sistema che erra nell'interpretazione dei dati di input, quindi non riconosce correttamente la situazione in cui è chiamato a intervenire, e produce un output difforme da quello che normalmente sarebbe richiesto, ovvero il sistema che pur interpretando correttamente i dati di input, produce un output incoerente rispetto all'interpretazione dell'input per un difetto di funzionamento interno (si pensi a un caso di *underfitting* o *overfitting*<sup>130</sup>).

---

di talché non ha senso vagliare la ragionevolezza algoritmica in via autonoma rispetto a quella dei programmatori.

<sup>130</sup> Per le definizioni si rinvia a quanto già illustrato nel cap. 1., par. 2.2.3.

Se l'errore del sistema produce o contribuisce a produrre un rischio che si traduce in un'offesa per uno o più beni giuridici, questa negligenza sarà causale rispetto alla produzione di una lesione o messa in pericolo di beni presidiati dalla legge penale. Si apre allora lo scenario della responsabilità penale del produttore, per aver prodotto e immesso nel mercato un bene che presenta dei difetti.

Consideriamo il caso del sistema che erra nell'interpretazione dei dati di input, quindi non riconosce correttamente la situazione in cui è chiamato a intervenire, e produce un output difforme da quello che normalmente sarebbe richiesto. Il problema che dobbiamo porci è anzitutto quali sono i fattori che consentono al sistema di *riconoscere* una data informazione e di *evitare* la produzione di un output che sia dannoso per un certo bene giuridico.

Come si è visto, riconoscibilità e evitabilità sono i parametri del giudizio probabilistico del fatto colposo utilizzati per la verifica ex ante dell'adeguatezza tra il comportamento realizzato e l'evento; possiamo allora trarre spunto dalle riflessioni della dottrina sul parametro della riconoscibilità per svilupparle sul terreno della responsabilità del produttore del sistema intelligente.

Abbiamo prima condiviso l'opinione di chi sostiene che l'*Erkennensollen* rilevante ai fini della tipicità colposa si compone di una base nomologica e di una ontologica che vengono costruite in rapporto a un agente modello.

Questo schema potrebbe essere riproposto per il manufatto intelligente che è tenuto a riconoscere la situazione di fatto per evitare di addivenire a una decisione che esponga a pericolo dei beni giuridici.

Dallo studio condotto nel primo capitolo si è ricavata una distinzione tra modelli di apprendimento simbolico e modelli di apprendimento connessionista. Ai primi, si ricorderà, appartengono gli alberi decisionali, mentre ai secondi si riconducono le architetture neurali. Questi modelli differiscono in punto di *rappresentazione della conoscenza* e di *acquisizione della conoscenza*. I modelli simbolisti adottano una rappresentazione della conoscenza che utilizza regole logiche seguendo lo schema top-down, mentre le architetture neurali hanno una rappresentazione della conoscenza che prescinde da schemi pre-conoscitivi di tipo logico, perché la rappresentazione della conoscenza è distribuita e avviene in modo parallelo attraverso la connessione tra i neuroni, secondo lo schema bottom-up. Quanto alle tecniche di acquisizione della conoscenza, i modelli simbolisti seguono un percorso consequenziale dettato dalle regole

logiche, mentre nei modelli connessionisti l'apprendimento è parallelo perché tutte le componenti della rete neurale apprendono simultaneamente.

Tracciando un parallelismo con lo schema della riconoscibilità quale parametro di valutazione ex ante dell'agire colposo, mi pare possibile assimilare la base nomologica e quella ontologica della riconoscibilità dell'agente umano rispettivamente alla rappresentazione e alla acquisizione della conoscenza in un sistema di IA. Infatti, le tecniche di acquisizione della conoscenza consentono al sistema di elaborare i dati della realtà esterna recepiti dai sensori, così come la base ontologica della riconoscibilità permette di recepire gli elementi di fatto di una situazione concreta, si tratta cioè dell'acquisizione della conoscenza del dato esterno. La rappresentazione della conoscenza nel modello di IA fa seguito all'acquisizione della conoscenza e serve a leggere, interpretare il dato acquisito attraverso un procedimento logico-simbolista o connessionista con rete neurale; allo stesso modo, il momento nomologico della riconoscibilità nel fatto colposo consiste nell'attivare le generalizzazioni causali note all'agente modello che gli consentono di pervenire alla rappresentazione dei rischi di una certa condotta.

Diremo, allora, che in un sistema di IA la base nomologica della riconoscibilità consiste nel *dovere* di possedere modelli di rappresentazione della conoscenza che permettono di individuare *correlazioni causali* tra input e output, mentre la base ontologica si sostanzia nel *dovere* di comprendere gli elementi della situazione di fatto in cui il sistema si trova a operare attraverso le tecniche di acquisizione della conoscenza e i sensori che rendono percepibile la realtà esterna.

La componente nomologica della riconoscibilità artificiale ci richiede di soffermarci ancora sulle caratteristiche dell'apprendimento artificiale. Come già osservato nel capitolo introduttivo, nella progettazione dei sistemi di IA si è abbandonata l'idea di costruire macchine pensanti, cervelli meccanici che emulino le capacità cognitive umane. Ad oggi, infatti, la scienza non è ancora in grado di spiegare come gli scambi tra i neuroni si traducano in "coscienza", di conseguenza la tecnica non è capace di realizzare un sistema dotato di un proprio pensiero e di una propria creatività<sup>131</sup>.

La ricerca si è allora indirizzata verso la creazione di modelli connessionisti, con l'obiettivo ben più modesto di creare macchine che non pensano, ma calcolano, fornendo

---

<sup>131</sup> S. Amato, *Biodiritto 4.0. Intelligenza artificiale e nuove tecnologie*, Torino, Giappichelli, 2020, p. 86.

una soluzione a partire dai dati che hanno a disposizione e dalle istruzioni comportamentali che vengono impartite nella fase di progettazione attraverso la tecnica dell'apprendimento supervisionato<sup>132</sup>.

L'assenza di una capacità di ragionamento della macchina comporta che essa non è in grado di comprendere il *significato* di un dato<sup>133</sup>, piuttosto essa si limita a instaurare una correlazione, associando il dato a un certo risultato che può consistere in un comportamento da tenere. Ad esempio: l'algoritmo impara a riconoscere il segnale di 'stop' attraverso i dati che vengono inseriti nel sistema e le indicazioni impartite dall'addestratore; ciò però non sarà sufficiente per comandare la guida della vettura autonoma: occorrerà altresì che all'algoritmo sia indicato di arrestarsi non appena riconosce il segnale in questione. Si instaura allora una correlazione tra l'input (segnale di 'stop') e l'output (comportamento di arresto del veicolo) e per ottenere un simile risultato l'algoritmo va 'nutrito' non solo con dati, ma anche con prassi di risposta comportamentale alla classificazione dell'input. L'arresto dinanzi al segnale di stop sarà così realizzato senza che il sistema comprenda le ragioni di tale comando comportamentale; esso non è in grado di riflettere sul fatto che il mancato arresto può comportare uno scontro con un'altra vettura, sviluppando un ragionamento argomentativo come fa un cervello umano: "non serve che pensi, basta che funzioni"<sup>134</sup>.

Per queste ragioni si è qui detto che la base nomologica della riconoscibilità artificiale contiene un insieme di correlazioni, cioè di legami tra input e output, senza però essere in grado di capire le motivazioni che portano a un certo risultato a partire dal dato iniziale. L'output non ha per la macchina un legame causale con l'input, quindi la riconoscibilità artificiale, a differenza di quella umana, non può essere costituita da generalizzazioni o leggi causali che spieghino il nesso di derivazione di un evento a partire da una data condotta<sup>135</sup>.

---

<sup>132</sup> Si veda il cap. I, par. 2.3.1.

<sup>133</sup> Come spiega bene L. Floridi, *La quarta rivoluzione. Come l'infosfera sta cambiando il mondo*, Milano, Raffaello Cortina Editore, 2017, p. 185.

<sup>134</sup> Espressione impiegata da S. Amato, *op. cit.*, p. 90, per descrivere il funzionamento di Google Deep Mind, sistema artificiale esperto nel gioco da tavolo AlphaGo che nei 2015 ha sconfitto il giocatore umano che deteneva il primato mondiale di vittorie. Tuttavia, a differenza di una mente umana, "Google Deep Mind non sa di giocare; ma "sa" che deve vincere e adotta le strategie migliori per farlo. Non serve che pensi, basta che funzioni".

<sup>135</sup> Cfr. J. R. Searle, *Is the Brain's Mind a Computer Program?*, in *Scientific American*, January 1990, pp. 26-32.

Su questi aspetti si rinvia anche alla successiva Sez. II sulla causalità.

Tutto questo ci porta a concludere che la ‘diligenza artificiale’ prescinde dalla comprensione del significato e delle ragioni delle istruzioni che vengono fornite all’algoritmo da parte dell’addestratore durante la fase di apprendimento. La diretta conseguenza di ciò è che la ‘mente pensante’ durante il funzionamento di un sistema di IA resta sempre quella umana, più precisamente, quella del *trainer* in fase di addestramento, mentre nel momento di utilizzo è quella dell’utente.

### 6.2.2. Negligenza artificiale

Queste premesse sulla struttura della riconoscibilità umana e artificiale di una situazione di fatto sono utili per lo studio delle patologie del funzionamento del sistema di IA che possa dar luogo a responsabilità. Possiamo quindi individuare due principali motivi che ostano a un perfetto apprendimento da parte del sistema di IA: i) difetti nel momento della rappresentazione della conoscenza (base nomologica dell’apprendimento automatico) e ii) difetti nel momento dell’acquisizione della conoscenza (base ontologica dell’apprendimento automatico), per poi accostarvi iii) situazioni ibride, che coinvolgono tanto l’una, quanto l’altra base dell’apprendimento automatico.

i) Mi pare che si possa ritenere che i difetti riscontrati nella fase di rappresentazione della conoscenza siano imputabili al produttore, perché tale capacità di rappresentazione dipende strettamente dalla programmazione del sistema e dalle impostazioni scelte al momento della progettazione. Si pensi ai casi, prima solo accennati, di overfitting e underfitting. Se nella fase di addestramento il sistema impara a riconoscere in maniera esatta e perfetta i dati di input fornendo un output che corrisponde perfettamente alle aspettative, il risultato non potrà dirsi soddisfacente, perché il modello non è in grado di generalizzare bene e produrrà degli errori se nutrito con dei dati diversi da quelli inseriti nella fase di addestramento. Nel caso di underfitting, il modello presenterà delle inesattezze già nella fase di addestramento, perché non è in grado di riconoscere i dati di input, e continuerà a errare anche quando si troverà a esaminare dati diversi da quelli impiegati al momento dell’apprendimento.

ii) Diverso è il caso dei difetti nel momento di acquisizione della conoscenza. Una acquisizione errata del dato esterno può dipendere ad esempio da un malfunzionamento

dei sensori causato da un difetto di produzione, imputabile al produttore, ovvero da una cattiva manutenzione imputabile all'utilizzatore<sup>136</sup>.

iii) Vi sono poi situazioni ibride o comunque di più difficile classificazione, in cui l'evento lesivo è il risultato di una commistione di limiti (e non difetti) del sistema riscontrabili nella fase ontologica e nomologica. Questo accade se immaginiamo che un'acquisizione della conoscenza incompleta (e non errata) possa essere determinata dalle peculiarità del caso concreto, laddove si presentino circostanze tali da rendere del tutto unica e imprevedibile la situazione di fatto. Si pensi al caso dell'improvviso attraversamento di un pedone in autostrada nottetempo non percepibile tempestivamente dal sensore. Ovvero al caso in cui un certo dato sia percepito dai sensori, quindi acquisito, ma il sistema non sia poi in grado di elaborarlo poiché il carattere unico della situazione la rendeva non prevedibile dal produttore, il quale non ha quindi 'addestrato' il sistema al riconoscimento di una situazione simile.

Per evitare difetti di questo tipo nella fase di rappresentazione e acquisizione della conoscenza si dovrebbe richiedere al produttore di procedere in sede di progettazione o addestramento alla descrizione di tutte le possibili situazioni di esposizione a rischio dei beni giuridici, obiettivo di difficile realizzazione, perché è impossibile predire i dettagli fattuali di ogni singola situazione in cui il sistema si troverà a operare<sup>137</sup>; detto altrimenti, è impossibile predire il futuro. Ciò del resto è in linea coll'idea per cui i 'segnali di allarme' che rendono riconoscibile un certo pericolo "vanno visti, non previsti; sono materia di rilevazione, non di preveggenza" e la riconoscibilità ha natura ricognitiva e non investigativa, in quanto l'agente non è tenuto a esplorare tutti i fattori che rendono astrattamente possibile l'offesa<sup>138</sup>.

L'inesigibilità di una simile richiesta rivolta al produttore emerge con evidenza se ci spostiamo dal settore della IA applicata alle automobili a quello della IA applicata alla scienza medica. Si pensi al caso in cui un nuovo virus si manifesti dopo la immissione

---

<sup>136</sup> Anche C. Piergallini, *Intelligenza artificiale, cit.*, a p. 1757, evidenzia che "non di rado, eventuali malfunzionamenti del sistema robotico, non ascrivibili al controllore, potrebbero pure discendere da malgoverno manutentivo.

<sup>137</sup> Come rilevato anche da A. D. Selbst, *Negligence and AI's Human Users*, in *Boston University Law Review*, 100, 1315 ss., (2020) a p. 1324, con riguardo ai rischi connessi alla guida autonoma: "Autonomous vehicles will face unexpected changes: detours from road construction, drivers who break traffic laws or stop very suddenly, or other drivers misapprehending what the automated vehicle itself will do and reacting badly. Each of these will be unique in some way—the timing, the type of stimulus—such that the machine cannot possibly be trained on all of them", p. 1324.

<sup>138</sup> F. Giunta, *Culpa, culpa, cit.*, p. 16.

nel mercato del robot deputato alla diagnosi di malattie infettive. Il virus non sarà riconosciuto dal sistema poiché non inserito tra le possibili infezioni e quindi non sarà diagnosticato. Per una simile limitatezza algoritmica non potrà certo muoversi un rimprovero per colpa al produttore visto che la scoperta del virus è successiva alla programmazione dell'algoritmo. Una colpa potrà invece riscontrarsi laddove, a seguito della scoperta del virus, il produttore non fornisca un aggiornamento del software che contempli la nuova patologia<sup>139</sup>.

Abbiamo così tracciato una prima distinzione tra situazioni di cui è possibile fornire *ex ante* una descrizione in termini di pericolosità, e situazioni in cui tale descrizione *ex ante* della pericolosità risulta impossibile, poiché correlate a dinamiche concrete difficilmente prevedibili e quindi non immaginate nella fase di addestramento dell'algoritmo e proposte come esempi del training set.

Va a questo punto altresì considerato che l'imprevedibilità del risultato di output può essere dovuta non tanto all'incapacità di condensare tutte le variabili concrete in una programmazione algoritmica, quanto piuttosto all'evoluzione dell'algoritmo per effetto dell'autoapprendimento durante la fase di utilizzo<sup>140</sup>. In molti casi, poi, i sistemi di IA interagiranno tra di loro attraverso le tecnologie di *cloud computing* che mette in contatto IA operanti in ambienti diversi e questo comporta un incremento esponenziale della conoscenza iniziale<sup>141</sup>. L'accrescimento della conoscenza potrà condurre a esiti decisionali diversi da quelli valutati in fase di test di validazione da parte del produttore, che sono pertanto imprevedibili<sup>142</sup>, col risultato che la decisione dell'algoritmo di ML non è sempre imputabile alla programmazione che ne ha preceduto l'immissione in commercio.

---

<sup>139</sup> K. Chagal-Feferkorn, *The Reasonable Algorithm*, cit., p. 136.

<sup>140</sup> F. P. Hubbard, "Sophisticated Robots": *Balancing Liability, Regulation, and Innovation*, 66 FLA. L. REV. (2014), 1803 ss., p. 1851, che utilizza l'espressione "emergent behavior".

<sup>141</sup> R. Borsari, *Intelligenza Artificiale e responsabilità penale: prime considerazioni*, in *MediaLaws*, p. 265.; F. P. Hubbard, cit., p. 1851, il quale nota come ciò comporta notevoli difficoltà probatorie in sede processuale.

<sup>142</sup> K. Chagal-Feferkorn, cit., p. 133, la quale sottolinea come la complessità aumenta se si considera che molti algoritmi di *machine learning* accrescono la loro conoscenza per effetto di interazioni con la rete, in quanto *online-based* (anche p. 135 sul punto). Osserva, inoltre, l'A., che il programmatore potrebbe sicuramente impostare dei limiti alla capacità di autoapprendimento, e inserire un meccanismo di selezione dei dati, così come sottoporre eventuali modifiche al processo decisionale dell'algoritmo alla sua pre-approvazione. Nondimeno, questa ultima caratteristica potrebbe rappresentare un impedimento che osterebbe alla funzionalità e utilità dell'algoritmo: si pensi alle ripercussioni che un simile sistema avrebbe sulle auto a guida autonoma, che smetterebbero di funzionare e di essere utilizzabili se ogni volta che siano messe dinanzi a situazioni nuove debbano attendere l'approvazione del programmatore (*ibidem*, p. 134).

Si può allora osservare che non sempre la ‘negligenza artificiale’ comporta in via automatica la responsabilità del produttore<sup>143</sup>. Piuttosto si potrebbero individuare tre situazioni-tipo: a) casi in cui il sistema doveva comportarsi in modo diligente e lo ha fatto; b) casi di negligenza artificiale in cui il sistema non ha agito secondo le aspettative; c) casi in cui il sistema non ha agito secondo le aspettative, ma non avrebbe potuto fare altrimenti, che potremmo definire di negligenza artificiale temperata o di inesigibile diligenza. Queste opzioni sono riprodotte nella tabella riassuntiva sotto riportata.

Caso a)	Produttore diligente	Algoritmo diligente
Caso b)	Produttore negligente	Algoritmo negligente
Caso c)	Produttore diligente (rischio consentito)	Algoritmo negligente (diligenza non esigibile)

Nella costellazione di casi di errore algoritmico rientrano le ultime due situazioni-tipo, ma gli esiti in punto di responsabilità saranno differenti. Se la negligenza artificiale dà luogo a responsabilità del produttore (ipotesi *sub b*)), non può dirsi parimenti per le ipotesi del terzo tipo, cioè per i casi in cui il produttore sulla base delle conoscenze tecnico-scientifiche disponibili non è in grado di ideare e addestrare l’algoritmo a fronteggiare nel migliore dei modi la situazione concreta in modo da evitare un certo rischio. È questo un margine di rischio che l’ordinamento tollera nel momento in cui autorizza un’attività di produzione di beni in cui il margine di errore pari a zero non esiste<sup>144</sup>, perché l’errore è intrinseco a sistemi decisionali che procedono per correlazioni e generalizzazioni a partire da un set di dati di input per produrre un output. Siamo quindi di fronte all’area del rischio consentito, nella quale il produttore agente non risponderà per colpa con riguardo ai danni “prevedibili, ma verificatisi nonostante la fedele osservanza di regole tecniche”<sup>145</sup>, e nella quale si colloca il punto di equilibrio tra le opposte esigenze di tutela dei beni giuridici minacciati e di svolgimento di attività utili ma intrinsecamente pericolose, individuato graduando l’intensità di tutela dei beni esposti

<sup>143</sup> R. Borsari, *Intelligenza Artificiale*, cit., p. 265.

<sup>144</sup> Come efficacemente notato da A. D. Selbst, *Negligence and AI’s Human Users*, cit., a p. 1331: “Because AI will not prevent all accidents, the promise of AI is to reduce—not eradicate—errors. Thus, when AI is used, there will still be some errors that result in harm”. Si veda anche Iria Giuffrida, *Liability for AI Decision-Making: Some Legal and Ethical Considerations*, 88 FORDHAM L. REV. 439 (2019), p. 422

<sup>145</sup> Così F. Mantovani, *voce*, cit., p. 311, sul tema del rischio consentito.

in ragione del loro valore<sup>146</sup>. A ritenere in questi casi il produttore responsabile, l'ordinamento entrerebbe in contraddizione con se stesso, autorizzando, da un lato, la produzione di beni di IA ad alto rischio e, dall'altro, imputando al produttore ogni conseguenza dannosa prevedibile (se si considera che i sistemi di IA basati sulle tecniche di ML non sono per loro natura sistemi a rischio zero) e evitabile solo astenendosi dall'attività rischiosa.

Certo, potrà darsi l'eventualità di un errore algoritmico evitabile se il produttore avesse disposto di maggiori risorse non solo economiche, ma anche in termini di dati, che avrebbero consentito un addestramento più completo, idoneo a inglobare anche la fenomenologia presentatasi nel caso concreto.

La questione ci induce a interrogarci sul parametro di misura della diligenza del produttore modello. Deve sul punto essere accolto il modello deontologico che orienta lo standard di diligenza individuando le regole di condotta sulla base del criterio oggettivo della prevedibilità e riconoscibilità alla luce della miglior scienza ed esperienza del momento storico in quello specifico settore<sup>147</sup>, quindi la tecnologia da adottare non è quella generalmente impiegata, bensì quella disponibile sul mercato, anche se sistematicamente inutilizzata<sup>148</sup> e anche se ciò comporta costi notevoli o persino la rinuncia a esercitare l'attività pericolosa allorché la rinuncia sia l'unico strumento in grado di minimizzare i pericoli<sup>149</sup>.

---

<sup>146</sup> G. Forti, *Colpa ed evento*, cit., p. 459 ss.

<sup>147</sup> Così, F. Mantovani, voce, cit., p. 306, il quale osserva che il richiamo a livelli inferiori di scienza ed esperienza, quali quelli dell'uomo normale o dell'*homo eiusdem professionis et condicionis*, confonde il piano oggettivo della colpa con quello soggettivo, cioè la funzione preventiva della regola cautelare col parametro dell'esigibilità dell'osservanza di essa da parte dell'agente. Altra dottrina invece sposa il modello deontologico, senza però rinunciare al parametro dell'*homo eiusdem professionis et condicionis* sul terreno del fatto tipico colposo (così G. Fiandaca, E. Musco, *Diritto penale*, cit., p. 586), che invece è da Mantovani invocato esclusivamente in sede di misura soggettiva della colpa. La minoritaria concezione socio-prasseologica fa principalmente capo a F. Giunta, *Illiceità e colpevolezza*, cit., pp. 173 ss. e 242-249; *Id.*, *La normatività*, cit., pp. 97-100, il quale sostituisce la prevedibilità parametrata sulla figura dell'agente modello con gli usi derivanti dall'esperienza collettiva, che garantisce una maggiore predeterminabilità e reperibilità del comportamento corretto da parte del soggetto. In giurisprudenza, la concezione è stata accolta dalla sent. Ct. cost. n. 312/1996, letta in chiave critica, tra gli altri, da G. Marinucci (*Innovazioni tecnologiche*, cit. p. 31) e da D. Castronuovo, *L'evoluzione teorica*, cit., p. 1631. Di quest'ultimo autore è da condividere la precisazione, fatta nello stesso scritto a p. 1629, secondo il quale il riferimento alla consuetudine è corretto nel caso di regole di diligenza, prudenza e perizia solo se riferito al momento genetico di formazione degli standard cautelari (fonte sociale della colpa generica) e non anche all'individuazione della cautela doverosa secondo il parametro "così fan tutti".

<sup>148</sup> Così, G. Marinucci, *Innovazioni tecnologiche*, cit., p. 32; D. Pulitanò, *Igiene e sicurezza del lavoro (tutela penale)* in *Dig. Disc. Pen. – Aggiornato 2000*, p. 396 ss.

<sup>149</sup> G. Marinucci, *Innovazioni tecnologiche*, cit., p. 46, il quale osserva come la stessa Cassazione avesse affermato che la rinuncia all'esercizio dell'attività componeva il dovere esigibile dai gestori delle dighe nei disastri del Vajont e di Stava.

Se quindi le conoscenze e le tecnologie disponibili in un dato momento storico non consentono di far fronte ai rischi che certi sistemi di IA presentano, non è da escludere che il legislatore possa valutare di impedire a monte una determinata produzione o sperimentazione. La soluzione può essere meno drastica per i limiti derivanti dalle generalizzazioni che caratterizzano la decisione algoritmica. In tale evenienza, ai limiti del sistema può ben sopperire una progettazione che preveda l'invio di segnali di allarme all'utente in caso di incapacità di funzionamento del sistema e consenta sempre all'utente di riprenderne il controllo quando in uso. Questi vincoli di produzione possono ricondurre il rischio connesso alle generalizzazioni entro soglie tollerabili grazie al controllo umano significativo che può esercitare l'utente.

Un altro aspetto merita un breve approfondimento, quello del rilievo che deve essere attribuito alle dimensioni dell'impresa nel momento in cui si delinea il dovere di diligenza.

La dottrina<sup>150</sup> ha evidenziato che le diverse dimensioni delle imprese rilevano al fine di stabilire la riconoscibilità del rischio; segnatamente, si può pretendere dalla grande impresa una conoscenza di livello maggiore e più aggiornata rispetto a quanto possa essere richiesto alla piccola impresa, per la quale ultima la riconoscibilità del rischio diviene doverosa nel momento in cui le conoscenze del settore sono più avanzate e diffuse. Ma resta fermo che a una misura della riconoscibilità del rischio regolabile in ragione delle dimensioni dell'impresa non corrisponde l'adattabilità delle precauzioni e degli standard di sicurezza, nel senso che non sarebbe giustificata una loro rimodulazione verso il basso quando si tratti della alla piccola azienda. Infatti, il rischio acclarato cui sono esposti beni giuridici diversi da quelli patrimoniali, quali la vita e l'integrità fisica, non tollera una modulazione degli standard di sicurezza in ragione di un'analisi costi-benefici e ciò vale sia per la sicurezza nei luoghi di lavoro<sup>151</sup>, che per la sicurezza del prodotto messo in commercio. L'epilogo di tale assunto argomentativo è già intuito dal lettore: la piccola azienda dovrà rinunciare a una certa produzione tecnologica se non sarà in grado di dotare il prodotto di meccanismi di sicurezza e di arresto di emergenza utili a fronteggiare un rischio che le è noto e che è tollerato fintantoché controllabile da parte dell'utente. In questo modo si ricostruisce il legame tra riconoscibilità e evitabilità del rischio. Una volta che il rischio è riconoscibile sulla base delle conoscenze di cui dispone il produttore che attinge al patrimonio scientifico diffuso in un certo momento storico,

---

<sup>150</sup> Ancora G. Marinucci, *Innovazioni tecnologiche, cit.*, pp. 51-52 e 54.

<sup>151</sup> È in questa prospettiva che analizza il problema G. Marinucci, *op. cit.*

esso dovrà essere evitato. Ciò è in linea con l'idea secondo cui i contenuti del dovere di diligenza dell'*homo eiusdem professionis et condicionis* vanno definiti sulla base dei principi etico-sociali che stanno alla base del nostro assetto costituzionale, quindi in modo coerente coi doveri di solidarietà politica, economia e sociale<sup>152</sup>

Anche sull'evitabilità potranno riflettersi le dimensioni dell'impresa. Un'impresa di caratura internazionale o che ha accesso a una mole notevole di dati o a dati esclusivi che le vengono trasmessi attraverso legami commerciali con altre aziende avrà maggiori possibilità di produrre sistemi di IA più sofisticati e quindi anche più pronti a fronteggiare situazioni del tutto peculiari grazie all'addestramento su un set di dati di input più ampio e variegato. Si pensi ad esempio a un algoritmo per il riconoscimento di patologie addestrato su dati che vengono forniti direttamente dal servizio sanitario nazionale al produttore aggiudicatario di una procedura di evidenza pubblica per la fornitura di un algoritmo per gli ospedali pubblici<sup>153</sup>. Qui l'evitabilità si traduce già nel dovere di addestrare l'algoritmo in modo da riconoscere i rischi messi in evidenza dai dati di input di cui il produttore dispone.

Al contrario, nel caso di un'impresa che non gode di rapporti contrattuali con aziende o enti pubblici in grado di fornirle dati che possano migliorare le prestazioni dell'algoritmo, la pretesa circa l'evitabilità del rischio si sostanzierà solo nel dovere di dotare il sistema di meccanismi di controllo *ex post*, in fase di utilizzo, non essendo possibile per l'impresa disporre di dati che consentano all'algoritmo di procedere *ex ante* a una valutazione più dettagliata a causa della penuria dei dati di input.

Il tema che sta sullo sfondo di queste brevi considerazioni sulla rilevanza delle dimensioni dell'impresa è se il costo economico che l'azienda è chiamata a sostenere per adeguarsi alle migliori misure di sicurezza possa avere un qualche spazio nella determinazione delle regole cautelari. La risposta che è stata data, secondo la quale i costi economici non possono incidere sugli standard di sicurezza dei prodotti con una determinazione al ribasso dello stesso, va però arricchita di un'ulteriore precisazione. La legge non potrà imporre indistintamente a tutte le imprese di adottare i sistemi di sicurezza

---

<sup>152</sup> G.P. Demuro, *Voce Homo eiusdem professionis et condicionis (profili storici)*, in *Ec. Dir., I Tematici, Il reato colposo*, Milano, Giuffrè, 2001, p. 607 ss., p.608.

<sup>153</sup> L'esempio è sollecitato dalle riflessioni di K. Chagal-Feferkorn, *cit.*, p. 138, la quale mette in evidenza la differenza tra aziende di grandi dimensioni che possono contare su una mole notevole di dati, come le aziende a partecipazione statale, o le aziende che hanno contatti con l'esercito, e aziende con capitali più esigui e banche dati meno vigorose rispetto a quelle pubbliche, per concludere che la colpa del produttore non va valutata solo sulla base dei limiti che presenta l'algoritmo, dovendosi tenere conto altresì delle caratteristiche dell'impresa.

più costosi; piuttosto, andrebbe fissata una soglia di tutela ritenuta indispensabile, immancabilmente estesa a tutte le imprese del settore, sopra la quale può rilevare il limite del costo eccessivo, nel senso che la loro adozione non può essere imposta ove risulti eccessivamente costosa per la categoria cui l'impresa appartiene<sup>154</sup>.

Naturalmente questo è un discorso che, finora, si sviluppa sul piano oggettivo della colpa e che serve a delineare il dovere di diligenza e a dotarlo già di una maggiore aderenza all'agente concreto attraverso l'individuazione di più sotto-categorie della figura del 'produttore modello', nella quale può quindi essere tracciata una distinzione tra imprese di grandi, medie e piccole dimensioni, che abbia un riflesso in punto di riconoscibilità e evitabilità del pericolo. Resta fermo che la sede in cui si assisterà a un giudizio ritagliato sulle esatte caratteristiche dell'agente concreto è però un'altra, cioè il momento dell'accertamento della colpevolezza colposa, in cui si potrà valutare se il produttore poteva conformarsi alla regola di comportamento che impone una produzione attuata con i più elevati standard tecnologici e un insieme di dati di addestramento che sia ampio, variegato e, al contempo, non discriminatorio.

Abbiamo quindi qui sottolineato come, nell'ottica della struttura del reato colposo, l'impossibilità di evitare la verifica del fatto rilevi già in sede di dovere obiettivo di diligenza se il comportamento richiesto al produttore per evitare la situazione di danno o di pericolo esula dal compendio di doveri delineato alla luce della *Maßfigur* di produttore immaginata al posto di quella dell'agente concreto. Del resto, la risposta che abbiamo immaginato da parte dell'ordinamento in questi casi sarebbe la stessa di quella che si registra allorché siano coinvolte altre figure professionali, come i medici: così come al medico non è imposto dall'ordinamento il *dovere* di salvare un malato terminale perché non esistono cure in grado di evitare l'esito infausto, allo stesso modo non può al produttore essere imposto il *dovere* di produrre un sistema che non realizzi rischi per i quali non esistono tecnologie in grado di evitarli.

---

<sup>154</sup> Si veda sul punto Corte Cost. n. 127/1990, in *Giur. cost.*, 1990, p. 718, nota di L. Violini; in *Foro it.*, 1991, I, c. 36, con nota di R. Fuzio; commento anche in D. Pulitanò, *Diritto penale, cit.*, p. 289, con la quale la consulta era stata investita della questione di legittimità costituzionale della disposizione sui limiti di accettabilità di emissioni atmosferiche che dovevano essere fissati dalla PA alla luce della migliore tecnologia disponibile, sempreché l'adozione delle misure di contenimento delle emissioni non comportasse costi eccessivi. La Corte, con una sentenza interpretativa di rigetto, ha chiarito che vi è un limite "ultimo assoluto e indefettibile rappresentato dalla tollerabilità per la salute umana e dell'ambiente in cui l'uomo vive", cosicché "il limite del costo eccessivo viene in causa soltanto quando quel limite sia stato rispettato". Pertanto la PA non potrebbe imporre il ricorso a tecnologie che riducano ulteriormente il livello di inquinamento "se queste risultino eccessivamente costose per la categoria cui l'impresa appartiene".

Ciò che è inesigibile già sul piano del *dover essere* da parte del produttore, ad esempio la previsione del singolo e peculiare caso concreto in fase di addestramento del sistema, può però lasciare spazio a una pretesa dell'ordinamento rivolta all'agente umano. Per questo bisogna ora analizzare la posizione dell'utente e il contenuto del dovere di diligenza a lui indirizzato, cosicché la tabella che abbiamo sopra disegnato possa essere arricchita di un'ulteriore colonna, quella che descrive la posizione dell'utente.

## **7. La posizione dell'utente. Diligenza umana e affidamento**

L'analisi dei doveri di diligenza e dei rispettivi titolari nel contesto dei danni da prodotto intelligente deve a questo punto rivolgere lo sguardo alla figura dell'utente.

Uno dei primi rilievi si concentra sul fatto che l'offesa che richiama il rimedio offerto dal diritto penale, ragionevolmente, si realizzerà durante l'utilizzo del sistema di IA da parte dell'utente. È allora proprio dal contesto di uso del prodotto intelligente che l'indagine sulla posizione dell'utente deve prendere le mosse, osservando come l'uso del prodotto presuppone l'instaurarsi di una "relazione" tra l'utente e il sistema di IA, in cui accanto a quella che qui abbiamo definito "diligenza artificiale" si accosta la diligenza umana dell'utente.

La relazione utente-sistema di IA potrebbe essere letta alla luce del principio di affidamento<sup>155</sup>, di cui il rapporto uomo-macchina ridisegnerebbe i poli, sino ad oggi rappresentati da individui, agenti umani<sup>156</sup>.

È certo che se escludiamo la responsabilità singolare del sistema, il principio di affidamento mette in relazione l'utente e il produttore, ma quest'ultimo viene chiamato in causa solo in via mediata, per effetto del rapporto che l'utente ha col prodotto intelligente.

È anche vero che, così impostata la relazione, il principio di affidamento potrebbe sussistere tra ogni consumatore e ogni produttore in ragione del legame creato dal bene

---

<sup>155</sup> Sul principio di affidamento, si veda la monografia di M. Mantovani, *Il principio di affidamento nella teoria del reato colposo*, Milano, Giuffrè, 1997.

<sup>156</sup> Un cenno all'affidamento che l'utente ripone sulla decisione del sistema di IA e i relativi riflessi sulla colpa è fatto da M. Di Florio, *Il diritto penale che verrà, cit.*, il quale sostiene che "la colpa dell'uomo potrebbe comunque ritenersi "lieve", se non del tutto esclusa, alla luce dell'affidamento riposto nelle maggiori capacità di carattere tecnico e calcolatorio della macchina", mentre "residuerrebbe, tuttavia, un margine di colpa "grave" quando l'errore dell'IA sia stato determinato da imperizia umana (ad es., per scarsa conoscenza di statistica) nell'interpretazione dei modelli previsionali", pp. 13-14.

acquistato dall'uno e prodotto dall'altro<sup>157</sup>. In questa sede, però, si vuole far leva sul bagaglio conoscitivo inserito nel sistema per verificare se un affidamento possa essere riposto sul prodotto intelligente, che possiede una sorta di “conoscenza” e dal quale ci si aspetta un certo livello di diligenza (artificiale) che è quella riconducibile a un prodotto modello della stessa specie, così che l'utente assuma delle scelte comportamentali sulla base di tale attesa<sup>158</sup>.

Quello della IA è un contesto in cui si assiste a una divisione di compiti tra utente e macchina, entrambi tenuti a conformarsi a norme astratte di comportamento vigenti nel settore di intervento (guida su strada, interventi medici, ecc...), per cui nella rispettiva area di competenza artificiale si crea una sfera di conoscenza algoritmica sulla quale l'utente può confidare, allo stesso modo in cui nell'attività medica di *équipe* la specializzazione di ciascun medico che compone la squadra giustifica il deficit conoscitivo degli altri medici, idoneo a creare un terreno favorevole all'applicazione del principio di affidamento<sup>159</sup>.

L'utente fa quindi affidamento sulla conoscenza derivante dai dati inseriti nel sistema -al pari di quanto accade nelle relazioni umane tra professionisti in cui il principio di affidamento trova spazio applicativo- e può fare affidamento sul fatto che il sistema artificiale agisca nel rispetto dello standard di diligenza artificiale promesso dal produttore. Certo, il paragone può reggere solo su un piano superficiale, perché nel rapporto tra agenti umani si fa affidamento sulla “conoscenza umana” che è diversa dalla

---

<sup>157</sup> E in effetti è all'interno della relazione produttore-consumatore che il principio di affidamento è stato studiato nell'ambito della responsabilità da prodotto. Più in particolare, è stato oggetto di studi l'affidamento che il produttore può riporre su un uso del prodotto conforme alle istruzioni da parte del consumatore (L. Kuhlen, *Fragen einer strafrechtlichen Produkthaftung*, Heidelberg, 1989, su cui v. M. Mantovani, *Il principio, cit.*, p. 160, nota 166).

<sup>158</sup> Secondo il Marinucci, il principio di affidamento è una conseguenza della pluralità di agenti modello cui corrisponde un certo dovere di diligenza, poiché i consociati nei rapporti di relazione orientano il loro comportamento confidando nel fatto che chi appartiene a una certa categoria di agente modello, cui a sua volta è ricondotto un dato tipo di diligenza, si comporterà secondo lo standard di diligenza che da quella categoria è preteso dall'ordinamento. G. Marinucci, *La colpa, cit.*, pp. 198 ss. Il principio di affidamento è allora correlato alla concezione normativa della colpa, in base alla quale in sede di *Tatbestand* si delinea una pretesa di diligenza che l'ordinamento rivolge a tipi di consociati e tali pretese fondano l'aspettativa giuridica in ordine al suo effettivo adempimento, M. Mantovani, *Il principio, cit.*, p. 58.

<sup>159</sup> I settori in cui la giurisprudenza, tedesca e italiana, hanno fatto maggiormente applicazione del principio in esame sono quello dell'attività medica di *équipe* e della circolazione stradale. Nel primo caso la sproporzione tra il sapere del singolo e quello degli altri operatori è di tipo qualitativo, nel senso che ciascun medico ha competenze diverse cosicché in certe aree bisognerà fare affidamento sulle conoscenze e competenze altrui; nel caso della circolazione stradale il deficit di conoscenza è invece di tipo quantitativo, dovuto al grande numero di utenti della strada che potranno tenere comportamenti diversi, cosicché si fa affidamento sul rispetto da parte di chi si mette su strada delle regole che regolano la circolazione. Sul punto ancora M. Mantovani, *Il principio, cit.*, p. 59.

“conoscenza artificiale”, in quanto solo l’uomo è un essere che agisce con coscienza e che è in grado di percepire tutte le sfaccettature del singolo caso<sup>160</sup>.

Il limite di una simile impostazione che porta a leggere il rapporto tra utente e sistema alla luce del principio di affidamento emerge con evidenza se si considera che il principio in esame fa leva su un’aspettativa comportamentale che, ancor prima di avere carattere sociale, è un’aspettativa giuridica, nel senso che lo standard comportamentale atteso è quello preteso dall’ordinamento per la categoria di attori cui appartiene l’agente in concreto<sup>161</sup>.

Il richiamo al principio di affidamento in questa sede non può pertanto essere sovrapposto del tutto all’utilizzo che dello stesso è fatto nel rapporto tra agenti umani e ciò per due ordini di ragioni. La prima, di tipo tecnico-informatico, che prende in considerazione il fatto che nel contesto dell’IA l’affidamento è sempre “attenuato” in ragione dei limiti che l’intelligenza non umana presenta. La seconda, di tipo più strettamente giuridico, diretta conseguenza della mancata elaborazione di una vera personalità giuridica dei sistemi di IA. Il risultato cui si perviene è che, allo stato attuale, il richiamo al principio di affidamento risulta improprio o comunque prematuro, poiché occorre prima che l’ordinamento vanti un’aspettativa comportamentale nei confronti proprio del sistema di IA. Finché ciò non si verificherà, il principio di affidamento varrà nella relazione che si instaura attraverso il prodotto intelligente tra l’utente e produttore, il quale -egli sì e non il sistema di IA- è riconducibile entro una categoria sociale dalla quale l’ordinamento si attende l’osservanza di determinate regole di diligenza fissate alla luce del parametro dell’agente modello.

Individuati allora i poli della relazione nella figura del produttore, da un lato, e in quella dell’utente, dall’altro, si dirà che il primo confida nell’utilizzo del prodotto in modo conforme alle istruzioni (in ciò consiste il dovere di diligenza dell’utente che fonda

---

<sup>160</sup> Va però dato atto dei tentativi di rendere i sistemi di apprendimento “consapevoli delle proprie competenze” e di dotarli di una progettualità che sia in grado di trasformare la macchina “da strumento passivo” dei bisogni umani, “in partner attivo”, come sottolineato da S. Amato, *cit.*, p. 89, il quale così descrive gli obiettivi del programma *Competency-Aware Machine Learning (CAML)* del DARPA, sul quale è possibile reperire maggiori informazioni al sito dell’Agenzia, <https://www.darpa.mil/program/competency-aware-machine-learning>, ove nella presentazione del progetto si legge: “*DARPA’s Competency-Aware Machine Learning (CAML) program addresses this challenge by enabling learning systems to be aware of their own competency. Systems will have knowledge of their learned abilities, the conditions under which those abilities were learned, knowledge of their resultant task strategies, and the situations for which those strategies are applicable*”.

<sup>161</sup> M. Mantovani, *cit.*, p. 135.

l'affidamento del produttore); il secondo confida nel fatto che il prodotto sia privo di difetti e sia stato realizzato nel rispetto degli standard previsti dalla normativa di settore.

Iniziamo allora a selezionare i doveri dell'utente modello, al quale si rivolgono anzitutto norme cautelari caratterizzate dal dovere di informarsi circa le condizioni di utilizzo del prodotto intelligente e le capacità dello stesso<sup>162</sup>, al fine di poter prevedere i casi in cui si manifestano i limiti della diligenza artificiale.

Se il dovere di informazione rappresenta il primo dei doveri di diligenza dell'utente-consumatore di un bene intelligente<sup>163</sup>, successivamente, durante l'utilizzo del sistema, il dovere di diligenza dell'utente consisterà in un dovere di vigilanza sul corretto utilizzo del manufatto, ma non si esaurisce in questo: come più volte ricordato, poiché i sistemi di IA 'debole' sono incapaci di elaborare forme di ragionamento simili a quelle dell'uomo, l'affidamento che l'utente ripone sul corretto funzionamento del sistema non è senza limiti. In presenza di circostanze concretamente percepibili che lascino presagire situazioni di pericolo non fronteggiabili dalla limitata diligenza artificiale, sarà richiesto l'intervento umano, con conseguente conversione del dovere di vigilanza in dovere di intervento, finalizzato a impedire l'integrazione di un'offesa penalmente rilevante.

## 7.1 Obblighi informativi dell'utente

Un ulteriore profilo da prendere in considerazione nell'analisi della posizione dell'utente è quello degli obblighi informativi e dell'eventuale responsabilità risultante dalla loro violazione<sup>164</sup>.

---

<sup>162</sup> Ma vanno condivise le perplessità espresse da T. Leiman, *Law and tech collide, cit.*, p. 261, in tema di auto a guida autonoma, secondo la quale le conoscenze fornite all'utente sul funzionamento dei sistemi di guida assistita saranno limitate, in quanto tali sistemi hanno un notevole valore commerciale e pertanto l'interesse dei produttori è quello di non rivelarne il funzionamento per evitare di rendere nota a terzi la tecnologia.

<sup>163</sup> Così anche G.E. Marchant, R.A. Lindor, *The Coming Collision Between Autonomous Vehicles and the Liability System*, 52 *Santa Clara l. Rev.* 1321, 1334 (2012), p. 1327.

<sup>164</sup> Secondo un certo orientamento dottrinale (N. Pisani, *La "colpa per assunzione" nel diritto penale del lavoro*, Jovene, Napoli, 2012; *Id.*, voce *Colpa per assunzione*, in *Ec. dir. I Tematici, Il reato colposo*, Milano, Iuffrè, 2021, pp. 233 ss.; R. Bartoli, *Colpevolezza: tra personalismo e prevenzione*, Torino, Giappichelli, 2005, p. 204; D. Piva, *Spunti per una riscoperta della colpa per assunzione*, in *DisCrimen*, 9.9.2020), la violazione degli obblighi informativi darebbe luogo a colpa per assunzione, istituto che viene in rilievo quando il soggetto ha intrapreso un'attività pericolosa ma giuridicamente autorizzata per l'espletamento della quale le regole cautelari violate richiedono specifiche capacità fisico-intellettuali o conoscenze tecniche di cui l'agente non disponeva, con conseguente realizzazione di un evento offensivo, sempreché l'agente conosceva o avrebbe dovuto conoscere la sua inadeguatezza rispetto allo svolgimento dell'attività da intraprendere (nozione tratta da E. Mezzetti, *Colpa per assunzione*, in *Scritti in memoria di Giuliano Marini*, Jovene, Napoli, 2010, p. 514). In particolare, dalla prospettiva del soggetto responsabile il comando suonerebbe sempre come "se non sei capace di tenere le condotte dovute, astieniti dall'intraprendere l'attività

La colpa da “omessa informazione” presuppone naturalmente l’esistenza di un apparato di regole in grado di guidare il comportamento dell’agente dinanzi alla situazione di pericolo. Orbene, nell’ambito delle attività di produzione e utilizzo di beni, questo apparato normativo può consistere nelle condizioni d’uso del prodotto, anche intelligente. La via è allora presto indicata per riflettere sul ruolo che l’istituto della colpa per violazione dei doveri di informazione ha nel panorama dei sistemi intelligenti.

Come ci insegna l’approccio europeo alla tematica dell’IA, ai fini regolatori è sempre bene distinguere tra sistemi ad alto e basso rischio, ed è per i primi che si aprono margini per una colpa dell’utente da omessa informazione. Questi, infatti, deve essere un utente responsabile, che si rende conto della complessità del sistema e che per questo non lascia all’improvvisazione l’utilizzo dello stesso. Un’indicazione in tal senso perviene ancora una volta dalla proposta di regolamento dell’aprile 2021 che, all’art. 29 c. 1, stabilisce che “gli utenti di sistemi di IA ad alto rischio usano tali sistemi conformemente alle istruzioni per l’uso che accompagnano i sistemi”. Infatti, i sistemi ad alto rischio devono sempre essere accompagnati da istruzioni per l’uso contenenti informazioni “concise, complete” e “chiare”, che siano “accessibili e comprensibili per gli utenti” (art. 13), in assenza delle quali sarà certamente riscontrabile una colpa del produttore perché il prodotto presenta un difetto da informazione<sup>165</sup>.

Ma a ben vedere, oltre che a questa disposizione specifica sull’obbligo di informazione dell’utente prevista dalla proposta di regolamento per i sistemi ad alto rischio, tale obbligo di informazione può essere ricondotto più genericamente ai doveri di solidarietà sociale di cui all’art. 2 Cost., considerato che l’utente che decide di utilizzare un sistema ad alto rischio crea un’area di potenziali effetti dannosi non solo per sé, ma

---

pericolosa”, ma la responsabilità deriva dalla violazione della regola cautelare e non sta nella violazione di un obbligo di astensione (R. Bartoli, *cit.*, p. 204), per cui quando l’obbligo di astenersi è calato nel contesto di attività rischiose, ma consentite, il rimprovero per colpa deve ritenersi più correttamente legato all’inosservanza di obblighi di informazione e controllo (N. Pisani, *cit.*, p. 109; D. Piva, *cit.*, p. 4). Di diverso avviso altra parte della dottrina, che sposa la concezione tradizionale di colpa per assunzione, incentrata sulla violazione di una regola cautelare che impone un obbligo di astensione (E. Mezzetti, *Colpa per assunzione*, in *Scritti in memoria di Giuliano Marini*, Jovene, Napoli, 2010, pp. 513 ss.; F. Antolisei, *Manuale di diritto penale, cit.*; G. Fiandaca, E. Musco, *Diritto penale, cit.*, p. 584; G. Forti, *Colpa ed evento, cit.*, p. 291; M. Romano, *Commentario, sub art. 43/105, cit.*, p. 469.). Tradizionalmente si afferma che la regola cautelare violata è una norma di prudenza, che impone di astenersi dall’assumere il rischio correlato allo svolgimento di un’attività della quale non si possedevano le competenze opportune (F. Giunta, *La normatività della colpa penale, cit.*, p. 86 ss., in particolare p. 525 e 531), per cui può dirsi che “la misura oggettiva della colpa per assunzione è rappresentata dalla contrarietà a regole di legittimazione ad agire; laddove, generalmente, le cautele, la cui inosservanza è alla base dell’illecito colposo, si atteggiano invece, più propriamente, a regole di azione (E. Mezzetti, *Colpa per assunzione, cit.*, p. 514).

<sup>165</sup> Sul difetto da informazione, si veda C. Piergallini, *Danno da prodotto, cit.*, p. 47-48.

anche per terzi che potrebbero subire un pregiudizio. In questa prospettiva costituzionale, quello informativo non è più da intendersi come onere, né tanto meno come obbligo, ma diventa un dovere, strumentale alla protezione dei diritti altrui, che si inquadra nel più ampio compendio dei doveri di solidarietà sociale<sup>166</sup>.

Il dovere di informazione dell'utente non è una novità nel panorama della responsabilità da prodotto; l'utente è messo in grado di rispettarlo nel momento in cui gli viene fornito dal produttore un manuale di istruzioni dove vengono indicati anche i potenziali rischi che si possono presentare durante l'utilizzo del bene. Se quindi tale dovere di informazione sussiste per tutti i prodotti, è però vero che nel caso di sistemi intelligenti ad alto rischio, a differenza che per altri prodotti, l'obbligo di informazione è certo più stringente in ragione della loro maggiore potenzialità lesiva<sup>167</sup>. Inoltre va considerato che tale dovere di informazione può essere ricordato all'utente anche mediante dei messaggi relativi all'aggiornamento del software inviati attraverso il sistema dalla casa produttrice. Il dovere informativo è quindi caratterizzato anche da un *dovere di aggiornamento* che si aggiunge al dovere iniziale di conoscere le condizioni d'uso del prodotto, data la capacità di questi sistemi intelligenti di 'evolversi'. Si tratta di una condizione che coinvolge soprattutto i sistemi di *machine learning*, i quali sono soggetti a mutamenti per effetto dell'autoapprendimento. Così, chi utilizza un prodotto intelligente, senza essersi prima informato sulle condizioni d'uso, sulla sorveglianza richiesta, sulla gestione di eventuali rischi che potranno verificarsi, su come arrestare la macchina in caso di emergenza, violerà il dovere di prudenza che sta a monte e precede quello che normalmente mira a conformare il comportamento che è effettivamente condizionale rispetto all'evento. L'utente sarà responsabile per non essersi informato sulle condizioni per l'utilizzo del sistema e per aver intrapreso un'attività di cui non sa padroneggiare i pericoli. In queste ipotesi è violato un obbligo di *fare*, cioè di informarsi, mentre l'obbligo di astensione sarebbe più propriamente riferibile al caso in cui l'utente si sia informato

---

<sup>166</sup> In questo stesso senso, N. Pisani, *La "colpa per assunzione"*, cit., p. 150, in particolare nota 3.

<sup>167</sup> In senso difforme T. Leiman, *Law and tech collide: foreseeability, reasonableness and advanced driver assistance systems*, *Policy and Society*, (2021), 40:2, 250-271, p. 262, con riferimento al caso in cui il dovere di informazione si traduca in un "extra training" per l'utente del prodotto intelligente rispetto all'utente di un prodotto non dotato di tali caratteristiche. In particolare, la riflessione è sviluppata con riguardo ai conducenti di auto a guida autonoma, per sostenere che richiedere agli stessi di sottoporsi a una pratica ulteriore rispetto a quella richiesta per il conseguimento della patente di guida non sarebbe ragionevole in quanto tale requisito non è imposto dalla legge; inoltre potrebbe condurre alla conseguenza per cui uno standard di diligenza maggiore sarebbe richiesto a chi decide di guidare auto con sistemi automatici, notoriamente più sicure, e non a chi guida veicoli tradizionali che espongono maggiormente a rischio i consociati.

sull'utilizzo del sistema e, pur consapevole di non essere in grado di maneggiare con la massima diligenza l'apparecchio intelligente, si sia comunque deciso ad utilizzarlo, in violazione di un obbligo che è -questo sì- di astensione, che impone cioè di non intraprendere una certa attività sapendo di non essere in grado di gestirne i rischi. In questo secondo caso, la consapevolezza dei propri limiti potrebbe qualificare la colpa come colpa con previsione.

La peculiarità della colpa per violazione di doveri informativi emerge al momento di individuare la condotta oggetto della valutazione di conformità al parametro cautelare, che non è la condotta che ha direttamente cagionato l'offesa, ma quella che vede il soggetto immettersi nell'esercizio di un'attività senza essersi informato. In ciò si addensa la differenza rispetto a un giudizio di colpa che ha invece ad oggetto la condotta che è immediatamente causa dell'offesa. E infatti, nella colpa basata sulla violazione di obblighi di informazione il giudizio della conformità della condotta alla regola cautelare retrocede al momento in cui il soggetto ha intrapreso l'attività, scelta che si lega in via *mediata* all'offesa, la quale è invece *direttamente* cagionata dalla condotta successiva. La colpa, quindi, è riferita al "pre-comportamento", ed è accertata con un percorso a ritroso<sup>168</sup> rivolto alla condotta che precede quella che l'offesa ha materialmente causato<sup>169</sup>. Il soggetto ha agito con un deficit di informazione che avrebbe *dovuto* (e in ciò sta il *Sollen*) e potuto (*Können*) colmare. La colpa per inosservanza di doveri informativi, allora, crea delle tensioni col "principio di coincidenza" (*Koinzidenzprinzip*) che anima il più generale principio di colpevolezza, chiedendo che l'elemento soggettivo del reato debba essere valutato in rapporto al fatto tipico e non a un momento antecedente<sup>170</sup>. La legge e

---

<sup>168</sup> N. Pisani, *La "colpa per assunzione"*, cit., p. 104.

<sup>169</sup> Non è questo un concetto limitato alla sola sfera della colpa da omessa informazione, ma che abbraccia istituti diversi, tutti rientranti nella macro-categoria della colpevolezza, per cui si è stata elaborata la categoria della pre-colpevolezza, per la quale l'indagine sull'elemento soggettivo viene anticipata a un momento antecedente a quello in cui la condotta è stata effettivamente tenuta: si pensi alla tesi che rintraccia l'imputabilità nell'*actio libera in causa* volgendo lo sguardo al momento in cui il soggetto si è posto nello stato di incapacità di intendere o di volere (N. Pisani, cit., p. 112 e pp. 117 ss.); o, ancora, all'errore inevitabile sul precetto penale, sul quale si rinvia a F. Palazzo, *Ignorantia legis: vecchi limiti e orizzonti nuovi della colpevolezza*, in *Riv. it. dir. proc. pen.*, 1988, 9020 ss., p. 958; ma anche ai casi in cui il soggetto agisce in una situazione di pericolo che integra gli estremi di una scusante come ad esempio lo stato di necessità, come aggiunge R. Bartoli, *Colpevolezza: tra personalismo e prevenzione*, Torino, Giappichelli, 2005, pp. 150 ss. In tutti questi casi l'episodio criminoso si svolge in due momenti: al momento  $t_1$ , il soggetto può essere ritenuto responsabile, o perché capace di intendere o di volere (quindi è imputabile), o perché avrebbe potuto evitare l'errore sulla legge penale o di far sorgere la situazione di pericolo; ma è solo al tempo  $t_2$ , che è integrato il fatto tipico, quando il soggetto non è imputabile, non conosce la legge penale né è in grado di conoscerla, oppure agisce in una situazione di pericolo che fonda lo stato di necessità, R. Bartoli, cit., p. 194.

<sup>170</sup> R. Bartoli, *Colpevolezza: tra personalismo e prevenzione*, Torino, Giappichelli, 2005, p. 190.

il principio di colpevolezza, pertanto, indirizzano lo sguardo dell'interprete verso il comportamento che realizza il fatto tipico, mentre la violazione del dovere di informarsi lo rivolge a ciò che accade prima della verifica del fatto tipico, cioè al pre-comportamento<sup>171</sup>. Il principio di coincidenza individua allora i margini oltre i quali il giudizio di colpa non può spingersi che coincidono col perimetro del fatto tipico: oltre questo limite il giudizio di colpevolezza rischia di tradursi in giudizio sulla condotta di vita e sulla personalità dell'agente, cioè in una colpevolezza d'autore<sup>172</sup>, che non può essere accolta e che rischia di rileggere il principio di colpevolezza in modo funzionale alle esigenze di tutela della società<sup>173</sup> senza tener conto dell'effettiva esistenza di una responsabilità personale del soggetto. Il fatto tipico funge allora da argine all'introduzione nel nostro ordinamento di una responsabilità da pre-comportamento<sup>174</sup>.

Alla luce di questa conclusione va tracciato il limite entro il quale il dovere informativo che grava sull'utente di un sistema intelligente può assumere rilevanza penale. Se, come detto, il diritto penale deve avere uno sguardo limitato entro il fatto tipico, la domanda cui dobbiamo rispondere è a quali condizioni il dovere di informazione dell'utente rientra nella condotta tipica. La risposta va rintracciata nella struttura del reato colposo e, più nello specifico, nella sua parte normativa: l'obbligo di informarsi sulle caratteristiche del sistema entra a far parte del dovere di diligenza e lo connota, rendendolo astrattamente personale, cioè riferito alla figura dell'utente modello. Richiedere l'osservanza di tale obbligo non significa, allora, anticipare il momento della condotta punibile a un comportamento estraneo al fatto tipico, cioè ad esempio al momento successivo a quello dell'acquisto o al momento precedente il suo primo utilizzo da parte dell'utente. L'obbligo di informazione, a rigore, dovrebbe sì essere adempiuto in questi momenti e sicuramente già quando il prodotto intelligente viene usato per la prima volta, ma esso ritorna a essere attuale ogni volta che il prodotto è utilizzato, nel senso che un utilizzo "non informato" di un bene intelligente classificato come ad alto rischio è sempre contrario a diligenza, cioè al dovere di usare il sistema in modo conforme alle istruzioni fornite dal produttore.

---

<sup>171</sup> N. Pisani, *voce Colpa per assunzione*, cit, p. 240.

<sup>172</sup> N. Pisani, *La "colpa per assunzione"*, cit., p. 114.

<sup>173</sup> In questo senso, R. Bartoli, *cit.*, p. 214.

<sup>174</sup> Va dato atto degli sforzi finalizzati a rendere compatibile col sotto-principio di coincidenza le ipotesi di pre-comportamento colpevole, che la dottrina ha classificato nel modo seguente: modello dell'anticipazione, modello della finzione e modello della eccezione. Per i relativi approfondimenti si rinvia a R. Bartoli, *cit.*, pp. 191 ss.

Questa regola cautelare, prevista -come si è detto- dall'art. 29 della proposta di regolamento, contiene a monte un obbligo positivo e a valle uno negativo, rispettivamente, l'obbligo di informarsi e, se inadempito, l'obbligo di astenersi dall'agire non informato.

Le conclusioni che se ne traggono sono, anzitutto, che nella descrizione della condotta dell'utente modello, la regola cautelare si compone di due momenti: un primo positivo, cui corrisponde l'obbligo di informarsi sulle caratteristiche del sistema; un secondo, negativo, consistente nel dovere di astenersi dall'utilizzo laddove l'agente non si sia informato o lo abbia fatto in modo parziale o inadeguato, ovvero dopo essersi correttamente informato, sia consapevole di non essere in grado di possedere le conoscenze o le abilità necessarie per il corretto utilizzo dei beni (si pensi al prodotto altamente tecnologico che richieda il possesso di conoscenze tecniche che vadano oltre la lettura delle istruzioni fornite dal produttore). La regola cautelare ha allora carattere composito, risultando dalla commistione di un dovere di diligenza, che impone un *facere*, con un successivo dovere di prudenza, che impone un *non agere* o un *aliud agere*.

La seconda conclusione qui raggiunta è che questo rimprovero per colpa si mantiene entro il perimetro del fatto tipico e non è riferibile a un pre-comportamento perché l'obbligo di informarsi è reso attuale da ogni singolo atto di utilizzo del sistema. Se l'acquirente di un prodotto intelligente non si informa sulle condizioni del suo utilizzo prima di impiegare il prodotto per la prima volta e l'offesa a un bene giuridicamente protetto si verifica solo qualche mese dopo il primo utilizzo, non si fa retrocedere il rimprovero al momento che ha preceduto il primo utilizzo, momento in cui normalmente l'obbligo di informazione è adempiuto. Piuttosto, va ritenuto che l'obbligo di informazione si rinnova ogni volta che l'utente si accinge ad utilizzare il bene, utilizzo che si presenta, allora, come situazione "attualizzante" che concretizza il dovere di diligenza informativa<sup>175</sup>, cosicché risulta annullata quella distanza temporale tra realizzazione del fatto lesivo e inadempimento del dovere di informazione. Tale dovere entra a far parte del fatto tipico e ne segna il momento iniziale, cioè il momento a partire

---

<sup>175</sup> Sulle "situazioni attualizzanti" il dovere di informazione e il relativo rapporto con la posizione di garanzia, si rinvia a N. Pisani, *La "colpa per assunzione"*, cit., pp. 150 ss., riferite alla posizione dell'imprenditore gravato da una posizione di garanzia generale sulla salute e sicurezza sui luoghi di lavoro. Come si vedrà nel capitolo dedicato alla causalità, non è pacifico che i doveri gravanti sull'imprenditore, ove disattesi, diano luogo a una responsabilità per omissione, dovendosi piuttosto considerare che egli disattende un obbligo che compone il dovere di diligenza e che, pertanto, la sua condotta omissiva confluisce nella componente omissiva della colpa; ne deriva che mentre ogni condotta colposa attiva nasce da un'omissione, non è per ciò solo che il fatto tipico va costruito in modo omissivo.

dal quale il contegno dell'agente ha incominciato ad acquisire connotati penalmente rilevanti poiché in contrasto con un dovere attuale di diligenza rimasto inadempito.

## 7.2 Dovere di vigilanza e dovere di intervento

Nel suo rapporto con il sistema di IA, l'utente si atteggia a "controllore umano", la cui supervisione sull'operato del sistema si presenta in modo duplice: da un lato come meccanismo di salvaguardia, proteso a impedire i danni derivanti dal cattivo funzionamento del sistema; dall'altro come centro catalizzatore di responsabilità, cioè come soggetto cui ascrivere eventuali danni evitabili<sup>176</sup>. A questa funzione duplice del controllo umano corrispondono due doveri di diligenza rivolti all'utente modello, segnatamente il dovere di vigilanza e quello di intervento; il loro adempimento definirà il modello virtuoso di atteggiarsi dell'utente e il meccanismo di salvaguardia dei beni giuridici, a dispetto della loro violazione che, se accertata, comporterà la catalizzazione della responsabilità sulla persona dell'utilizzatore<sup>177</sup>.

Iniziando dall'esame del dovere di vigilanza, condizione attualizzante dello stesso è l'utilizzo del sistema da parte dell'utente. Se in genere l'uomo nel suo operato è agevolato dall'utilizzo di macchine, nel contesto della IA la prospettiva è rovesciata, ma resta comunque antropocentrica: la macchina non si limita più a coadiuvare l'uomo nelle sue attività, ma svolge essa stessa una certa funzione al posto dell'uomo, il cui compito si restringe alla supervisione dell'agente artificiale.

Il controllo umano sembra al momento imprescindibile e in questo senso è orientata anche la visione europea della IA. Un simile approccio non può che essere condiviso<sup>178</sup>,

---

<sup>176</sup> C. Piergallini, *Intelligenza artificiale, cit.*, pp. 1757-58.

<sup>177</sup> A. D. Selbst, *Negligence and AI's Human Users*, in *100 Boston University Law Review*, 2020, 1315 ss., osserva a p. 1345: "The human-in-the-loop aspect of the technologies that still rely on negligence law ensure that this type of wildly unexpected AI injury cannot happen, or in fact, the human would be reasonably blamed for it". L'A. commenta l'interessante esempio riportato da Lemley, Mark A. and Casey, Bryan (2019) "Remedies for Robots," *University of Chicago Law Review*: Vol. 86: Iss. 5, Article 3, p. 1311 ss., di un drone addestrato per raggiungere il centro di un cerchio. Il drone, dopo i primi tentativi nel corso dei quali riceveva rinforzi positivi per i successi nel compito assegnatoli, iniziava ad avere un comportamento diverso: giunto in prossimità del bordo del cerchio, improvvisamente vi si allontanava. Gli addestratori, allora, lo spegnevano e lo riponevano al centro del cerchio, per ricominciare l'esperimento. Dopo varie investigazioni, i programmatori scoprirono che il drone aveva capito che se si allontanava dal bordo, si sarebbe in qualche modo ritrovato "teletrasportato" al centro del cerchio, raggiungendo così il suo obiettivo. Selbst, *cit.*, p. 1345, allora osserva che se il funzionamento del drone fosse stato supervisionato da un essere umano, non appena esso iniziava ad allontanarsi dal cerchio l'uomo avrebbe potuto riprenderne il controllo e evitare che si distanziasse dal centro.

<sup>178</sup> Ma va dato atto delle osservazioni che la dottrina americana fa sul punto, allorché sottolinea come in certe situazioni l'esigenza di un intervento umano potrebbe creare rischi ulteriori rispetto a quelli che

anche alla luce del fatto che i sistemi di intelligenza artificiale debole hanno una cognizione computazionale limitata: essi si muovono bene entro un dato perimetro di conoscenza, ma non vanno oltre il campo conoscitivo tematico predisposto in fase di produzione<sup>179</sup>. Allo stato, esistono forme altamente intelligenti su temi delimitati e non è stata ancora creata una *Artificial general intelligence* (AGI) -anche detta intelligenza artificiale forte-, cioè un'intelligenza artificiale generale, di livello pari a quello umano<sup>180</sup>.

Il risultato che ne traiamo in termini di doveri dell'utente è che quello di vigilanza è un dovere reso attuale al momento dell'avvio del sistema e che permane durante tutta la fase di utilizzo dello stesso.

Il dovere di vigilanza è però suscettibile di mutare fisionomia, per tradursi in dovere di intervento. Al fine di individuare le circostanze in presenza delle quali il dovere di vigilanza dell'utente si tramuta in dovere di intervento possiamo ricorrere a degli esempi che rendano la nostra riflessione più agevole.

Pensiamo dapprima a un sistema intelligente utilizzato in ambito medico per riconoscere solo un numero limitato di patologie. La diagnosi riferita a una di quelle patologie sarà sicuramente più precisa e giungerà probabilmente anche in un tempo più celere rispetto a quanto non sappia fare un medico umano; nondimeno, il sistema sarà "intelligente" solo all'interno di quel campo (ad esempio nella diagnosi di patologie tumorali dell'occhio), ma una volta escluso che una di quelle patologie ricorra, non possiamo per ciò solo ritenere il paziente sano (la macchina non saprà riconoscere un'infezione sintomatica di una patologia a livello neurologico poiché quel tipo di malattia non rientra nella sua "conoscenza").

---

deriverebbero da un processo di funzionamento totalmente automatizzato che non richiede affatto la supervisione umana. L'esempio è legato alla presenza del sistema di direzione automatico del veicolo, per cui si osserva che una volta che il guidatore percepisce la concreta possibilità di una collisione imminente, potrebbe decidere di riprendere il controllo dello sterzo intervenendo però, a causa della situazione di panico, in un modo che aggrava la situazione. Si conclude, allora, che "*At some point, then, removing the human entirely from active driving may be safer than managing the "mushy middle" of shared human-machine operation*". B. W. Smith, *Automated Driving and Product Liability*, *Michigan State Law Review* 2017, no. 1 (2017), pp. 1-74, p. 49.

<sup>179</sup> E ciò vale anche nel caso di autoapprendimento: l'algoritmo accrescerà la sua conoscenza ma sempre al fine di realizzare il compito che gli è stato affidato. Se questo consiste nel riconoscimento di malattie tumorali dell'occhio, l'ampliamento della conoscenza non può aversi verso malattie neuronali per riconoscere le quali la macchina non è stata progettata.

<sup>180</sup> Molto chiaro sul punto già Jhon R. Searle, *Is the Brain's Mind a Computer Program?*, in *Scientific American*, January 1990, pp. 26-32; così anche Selbst, A. D., *Negligence and AI's Human Users*, cit., p. 1344; Erik Sofge, *Artificial Intelligence Will Not Obliterate Humanity*, *Popular Sci.*, Mar. 19, 2015, p. 36 ss.

Immaginiamo ora un altro scenario, in cui un comportamento diligente è tenuto in adempimento di una regola precauzionale di fonte sociale che fonda la colpa generica o, nel caso di regole scritte, di una regola elastica, che faccia rinvio a regole di comune esperienza, ora per la determinazione del comportamento diligente, ora per l'individuazione delle condizioni fattuali di operatività della regola<sup>181</sup>. La propensione del sistema di IA a eseguire delle istruzioni, la sua inabilità nello svolgere delle valutazioni e la sua assenza di capacità critica, comportano che un addestramento basato su una regola generica potrà produrre output poco chiari per il sistema, il quale non sarà in grado di estrapolare una regola modale di condotta valida per il caso concreto.

Prendiamo come esempio la regola “guidare con prudenza” desunta dal principio informatore della circolazione di cui all’art. 140 cod. strada, a mente del quale “*gli utenti della strada devono comportarsi in modo da non costituire pericolo o intralcio per la circolazione e in modo che sia in ogni caso salvaguardata la sicurezza stradale*”<sup>182</sup>. In questo caso la norma giuridica è tale solo apparentemente, poiché in realtà il suo contenuto è determinato attraverso il richiamo a norme cautelari sociali<sup>183</sup>. Proprio la sua essenza di norma sociale, e non giuridica, rende la regola comprensibile a un agente umano, il quale è in grado di declinarla e contestualizzarla nel caso concreto poiché è dotato di capacità critica. Una simile regola sarà, ad esempio, declinata nel dettato seguente: “in presenza di un semaforo verde, assicurarsi comunque che non vi siano auto provenienti lateralmente prima di proseguire la marcia”, o ancora: “in presenza di segnali stradali coperti con adesivi o graffiti, rallentare la marcia per assicurarsi di averne compreso il contenuto” . La capacità di ragionamento della mente umana permette di desumere da una regola generica o elastica (guidare con prudenza) una regola peculiare in un dato contesto (verificare che non vi siano auto nell’incrocio anche quando si ha il diritto di precedenza; rallentare in presenza di segnale stradale coperto da adesivi o graffiti al fine di verificarne il contenuto).

---

<sup>181</sup> Si tratta delle norme cautelari scritte che F. Giunta contrappone a quelle sufficientemente determinate, in *La normatività*, cit., p. 92, citando come esempio proprio l’art. 140 cod. strada che sarà a breve indicato anche qui come esempio.

<sup>182</sup> Norma che, secondo Marinucci, contiene una clausola di riserva esplicita -e non implicita come avviene nella più larga parte dei casi- che esprime l’idea per cui l’osservanza delle regole cautelari rigide è dovuta se da essa non derivi un aumento del rischio per i beni giuridici che la norma si prefigge di tutelare. L’art. 140 c. strada (previgente art. 101 all’epoca in cui scrive il Marinucci) è norma generale che limita l’operatività delle disposizioni del cod. strada, subordinandone l’applicazione al presupposto che la loro osservanza serva a eliminare i pericoli e non a crearne degli altri, *La colpa*, cit., pp. 248-249.

<sup>183</sup> F. Giunta, cit., p. 92.

Caliamoci ora nella prospettiva dell'agente artificiale. Ebbene, una regola del tipo "guidare con prudenza" sarebbe una regola "muta", dalla quale il sistema non è da solo in grado di estrapolare delle generalizzazioni da tradurre in regole adatte al contesto concreto che si viene a presentare; ciò vuol dire che il sistema non è capace di leggere la situazione fattuale di potenziale pericolo alla luce della generale regola "guidare con prudenza". Se, infatti, l'apprendimento umano può essere del tipo *top-down*, quindi conosciuto il concetto generale di prudenza è in grado di calarlo nel contesto concreto e ricavare una regola di condotta, ciò non vale per l'algoritmo a base neuronale. Il concetto di prudenza sarà "compreso" (sempreché di comprensione possa parlarsi con riguardo alle macchine) alla luce di esempi secondo lo schema conoscitivo *bottom-up*, che richiede allora che l'uomo addestri l'algoritmo a fronteggiare determinate circostanze con una certa risposta attitudinale. Tuttavia, se il caso concreto non rientra tra quelli con i quali l'algoritmo è stato addestrato, la macchina procederà con la generalizzazione, ma non è escluso che l'algoritmo generalizzi 'male' offrendo un output che cagioni un'offesa a un bene giuridico.

Lo stesso dicasi per lo schema conoscitivo *top-down* dei sistemi di IA con modello di apprendimento logico: il concetto di 'prudenza' sarà appreso dal sistema previa trasposizione da parte degli informatici di un processo logico (ad esempio quello degli alberi decisionali) che guidano l'algoritmo nell'addivenire a un dato output; può però sempre presentarsi un caso non conosciuto prima dal modello, al quale dunque non è già nota la risposta esatta di reazione a fronte di quelle condizioni fattuali, per cui sarà tenuto a ricorrere a delle generalizzazioni, con possibile margine di errore.

Le due situazioni appena illustrate sono esempi di cognizione algoritmica limitata vuoi perché l'algoritmo si confronta con una fenomenologia concreta che esula dal suo campo di conoscenza (è il caso della malattia all'occhio e del virus di nuova scoperta), vuoi perché a evitare l'esposizione a pericolo del bene giuridico e la conseguente offesa interviene una regola di diligenza dai confini mobili (si tratta di quei casi in cui verrebbe normalmente in gioco una colpa generica ovvero una colpa specifica in cui il comportamento diligente è descritto da una regola elastica, v. art. 140 c. strada), che mentre è comprensibile per un conducente umano, lo sarà con molte più difficoltà per un conducente artificiale.

Tali fenomenologie rientrano in quelle situazioni-tipo che nel trattare la posizione del produttore abbiamo classificato come ipotesi in cui il sistema non ha prodotto il

risultato atteso, ma non avrebbe potuto fare altrimenti (caso *sub c*)), situazione-tipo in cui, secondo il ragionamento proposto, non pare sostenibile una responsabilità del produttore, la cui azione si collocherebbe nell'area del rischio consentito dal legislatore (salve le precisazioni sviluppate *supra*). Ora possiamo aggiungere che, in questo spazio di rischio consentito per il produttore, può trovare posto la responsabilità dell'utente, a condizione che le circostanze del caso concreto siano tali da tramutare il dovere di vigilanza dell'utente sul sistema di IA in dovere di intervento che, se disatteso, pone le basi per una sua responsabilità a titolo di colpa. In presenza di quali condizioni sussiste il dovere di vigilanza dell'uomo è la questione che ora andremo a trattare.

Iniziamo dall'esame delle situazioni che esulano dal campo di conoscenza dell'algorithmico, riprendendo l'esempio delle patologie mediche non riconosciute dal robot-medico. Ebbene, se emerge in tali situazioni la limitatezza dell'algorithmo, è però vero che un'indagine incrociata più aderente al caso concreto potrà essere effettuata dall'uomo, il quale dispone di una conoscenza estesa a più patologie e non solo a quelle tumorali, anche se nell'ambito delle malattie tumorali le informazioni di cui dispone sono inferiori a quelle del medico robot immaginato nel nostro esempio<sup>184</sup>. In questa situazione il dovere di intervento è reso attuale dalla cognizione computazionale limitata dell'algorithmo, progettato per rispondere a compiti (*tasks*) specifici (riconoscimento dei tumori dell'occhio), per cui l'utente è chiamato a intervenire se nel caso concreto bisogna raggiungere scopi che esulano dalla progettazione dell'algorithmo (riconoscimento di malattie neuronali).

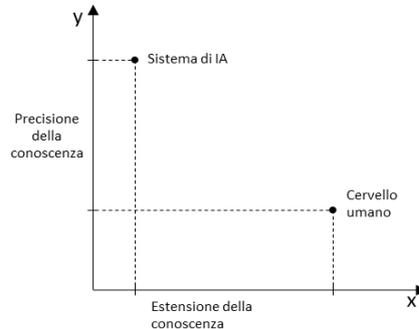
Trasponendo il confronto tra la conoscenza artificiale e quella umana in un diagramma cartesiano, in cui l'ascissa indica l'estensione della conoscenza e l'ordinata la precisione della conoscenza, collocheremo il "cervello" robotico nel punto più alto delle ordinate, ma a poca distanza dal punto zero nell'asse delle ascisse. La posizione del

---

<sup>184</sup> Un altro esempio dal quale emerge la differenza tra intelligenza umana e 'intelligenza' artificiale, riportato da K. Chagal-Feferkorn, *The Reasonable Algorithm, cit.*, p. 137-138, è quello di una donna che si reca al pronto soccorso accompagnata dal marito per delle lesioni che dichiara essersi procurata scontrandosi contro la porta. Il medico, durante la descrizione dell'accaduto da parte della donna, nota un atteggiamento introverso della stessa, la quale evita di incrociare lo sguardo del marito; rileva altresì un atteggiamento poco empatico dell'uomo e ciò gli fa sorgere il sospetto che si tratti di un caso di violenza domestica per il quale è necessario l'intervento di un assistente sociale come indicato dai protocolli dell'azienda sanitaria.

Se lo stesso caso fosse presentato a un medico-robot, dotato di innumerevoli capacità computazionali, ancorché dotato della capacità di desumere gli stati psicologici dal tono di voce, potrebbe non elaborare l'ipotesi dell'eventuale violenza domestica, così non attivando la procedura adeguata per il trattamento del caso.

cervello umano sarà invece all'opposto: essendo più estesa la sua conoscenza si troverà in un punto lontano dell'asse dell'ascissa, ma meno distante dallo zero nell'asse dell'ordinata.



Il dovere di intervento dell'uomo è allora una diretta conseguenza del limitato campo di conoscenze dell'algoritmo, che opera, e anche meglio dell'uomo, solo entro un ambiente cognitivo ristretto<sup>185</sup>: queste macchine 'intelligenti' vengono concepite per reagire alle situazioni previste da chi le crea, ma non già a quelle che sfuggono alla cognizione del programmatore<sup>186</sup>.

Passiamo ora all'altra situazione-tipo di conoscenza algoritmica limitata, quella in cui la regola cautelare elastica presenti un contenuto non comprensibile, cioè si caratterizzi attraverso il richiamo a una regola di fonte sociale che il sistema non è in grado di recepire. Ancora una volta si nota che questi limiti caratterizzano la sola IA e non anche l'agente umano, il quale è parte cosciente di quella società che elabora le regole sociali non scritte attraverso generalizzazioni, che costituiscono la fonte della colpa generica e integrano le regole elastiche fonti di colpa specifica.

Basti pensare, ad esempio, alla regola "verificare che non vi siano auto nell'incrocio anche quando si ha il diritto di precedenza" che, sebbene non codificata, è comunque desumibile dal guidatore a partire dalla più generale regola "guidare con prudenza";

---

<sup>185</sup> Come precisa Floridi, *op. cit.*, p. 155, "i sistemi di intelligenza artificiale più efficienti sono quelli che operano all'interno di un ambiente che è conformato intorno ai loro limiti. [...] Occorre adattare l'ambiente al robot per essere certi che questo possa operare in esso con successo" e osserva che "la vera difficoltà [per il sistema di IA ndr] p avere a che fare con l'imprevedibilità del mondo là fuori [...]". Ciò è conosciuto come *problema del frame*, relativo al modo in cui un agente situato in un contesto può rappresentarsi un ambiente che cambia e interagisce con esso nel tempo in maniera efficiente", per concludere che (p. 163) i sistemi di IA "non stanno diventando più intelligenti, rendendoci al contempo più stupidi. È il mondo invece che sta divenendo un'infosfera sempre più adatta alle [loro] limitate capacità".

<sup>186</sup> C. Piergallini, *Intelligenza artificiale, cit.*, p. 1759.

pertanto, l'aspettativa dell'ordinamento in ordine al dovere di diligenza dell'utente non si arresta al rispetto delle regole sul traffico regolato da semafori o da segnali di precedenza, ma si estende anche all'osservanza del più generale principio informatore della circolazione di cui all'art. 140 cod. strada che, nel caso concreto, richiede di verificare che ci siano le condizioni ideali per proseguire la marcia pur avendo un diritto di precedenza. È comunque vero che i sistemi intelligenti possono essere programmati per accertarsi che l'incrocio sia libero, anche quando il semaforo indichi la luce verde, ma deve esserci una programmazione apposita in tal senso, per cui non basterà indicare al sistema la regola "guidare con prudenza", ma bisognerà istruirlo in modo specifico, insegnandoli che "anche in presenza di un semaforo verde bisogna accertarsi che nessuno provenga dalle altre direzioni". Emerge, allora, il limite di cui già sopra si è detto, cioè che il produttore dovrebbe programmare tutte le possibili situazioni di pericolo che possono essere evitate osservando la regola "guidare con prudenza", pre-programmazione impossibile perché innumerevoli saranno i pericoli cui ci si espone mettendosi alla guida.

Peculiare è l'ipotesi del segnale stradale ricoperto con degli adesivi o graffiti. Un gruppo di ricercatori ha dimostrato che simili elementi di disturbo sono in grado di indurre in errore l'algoritmo, il quale interpreta un segnale di 'stop' ricoperto da adesivi come un segnale di limite di velocità pari a 45 km/h<sup>187</sup>. Anche qui, allora, la regola "guidare con prudenza" fallisce per l'algoritmo, che non percepisce l'adesivo e erra nell'interpretazione del segnale stradale e non riesce a concretizzarne il contenuto nella situazione concreta.

E queste sono solo alcune delle ipotesi in cui potrebbe configurarsi un dovere di intervento in capo all'utente.

A sostegno di un dovere di intervento dell'utente negli scenari in cui la capacità cognitiva del sistema di IA è circoscritta, va altresì aggiunto che nella fenomenologia concreta il dovere di diligenza dell'utente potrà arricchirsi di eventuali conoscenze di cui lo stesso disponga. Si pensi al caso -spesso richiamato dalla dottrina in materia di standard obiettivo del dovere di diligenza- del guidatore che percorre con frequenza un tratto di strada e che sa che molti incidenti si sono verificati in prossimità di quell'incrocio, ad esempio a causa di un guasto alla segnaletica luminosa che omette di far scattare in una delle due strade il segnale rosso che impone l'arresto. Si tratta di una regola di diligenza

---

<sup>187</sup> K. Eykholt et al., *Robust Physical-World Attacks on Deep Learning Visual Classification*, 2018, *arXiv.it*

che l'agente umano apprende grazie all'esperienza concreta e che col tempo andrà a costituire quelle superiori conoscenze causali che la dottrina ritiene di poter prendere in considerazione nel momento dell'individualizzazione dello standard obiettivo di diligenza<sup>188</sup>.

Siffatta individualizzazione è impossibile nei riguardi di un sistema di IA, che non è dotato delle capacità di una mente umana, non è un 'motore semantico', né un'intelligenza forte, ma piuttosto un'intelligenza debole tipica di un 'motore sintattico', e per questo non riesce a comprendere le sfumature della situazione concreta così come riesce a fare un agente umano<sup>189</sup>.

Questa differenza che permane tuttora tra agente umano e agente artificiale, dovuta proprio alle caratteristiche dell'IA leggera o debole, incide sul riparto delle responsabilità tra produttore e utilizzatore.

È quindi giunto il momento di arricchire di un'ulteriore colonna la tabella sopra presentata, in cui sia presa in considerazione la posizione dell'utente, segnalando il contenuto del suo dovere di diligenza. Si noterà che i doveri di informazione, vigilanza e intervento sono sempre presenti, ma resta fermo il discorso qui sviluppato circa la situazione che attualizza tali doveri, i quali restano latenti finché non si presentino circostanze di fatto in grado di attivarli. Segnatamente, è l'utilizzo del sistema a rendere attuale i doveri di informazione e vigilanza, mentre sono situazioni peculiari come quelle che si è qui cercato di individuare in degli esempi a attivare il dovere di intervento.

Caso a)	Produttore diligente	Algoritmo diligente	Utente diligente (dovere di informazione, vigilanza e intervento)
Caso b)	Produttore negligente	Algoritmo negligente	Dovere di vigilanza e intervento dell'utente
Caso c)	Produttore diligente (rischio consentito)	Algoritmo negligente (diligenza non esigibile)	Dovere di vigilanza e intervento dell'utente

<sup>188</sup> Sulla rilevanza delle speciali conoscenze "causali" ai fini dell'innalzamento della misura oggettiva di diligenza, M. Romano, *Commentario sistematico del codice penale*, Milano, Giuffrè, 2004, sub art. 43, p. 459; G. Fiandaca, E. Musco, *Diritto penale, P.te gen.*, Bologna, Zanichelli, 2019, p. 585; G. Forti, *Colpa ed evento nel diritto penale*, Milano, Giuffrè, 1990, pp. 272 ss.

<sup>189</sup> È qui ripresa la terminologia coniata da L. Floridi, *La quarta rivoluzione*, cit., p. 185, allorché si utilizzano le espressioni "motore semantico" e "motore sintattico" per indicare, rispettivamente, l'uomo e la macchina. Ad affermare che "computer programs are formal (syntactic)" e che "Human minds have mental contents (semantics)" fu Jhon R. Searle, *Is the Brain's Mind a Computer Program?*, in *Scientific American*, January 1990, pp. 26-32.

### 7.2.1 Macchine etiche? La violazione necessitata della regola di diligenza

Potremmo poi individuare un'altra circostanza in cui è reso necessario un intervento dell'utente durante l'utilizzo del sistema. Restiamo ancora nell'ambito della circolazione stradale: l'inserimento delle regole del codice della strada nel software di pilota automatico comporta la realizzazione di macchine intelligenti, che conoscono le regole di comportamento, alle quali si attengono con una maggiore aderenza rispetto a quanto non faccia un agente umano, il quale, ad esempio, potrebbe non esitare più tanto a rispondere a una telefonata mentre è alla guida.

Ma la scrupolosa osservanza di regole cautelari precostituite potrà veramente evitare forme di responsabilità colposa del produttore (e anche dell'utilizzatore)?

La risposta può essere ricostruita prendendo le mosse dalle caratteristiche delle norme precauzionali a struttura rigida. Queste non si prestano a essere influenzate dalle peculiarità del caso concreto e aspirano a realizzare una funzione di prevenzione e di protezione dell'integrità di beni giuridici più avanzata e intensa rispetto a quanto facciano gli schemi elastici<sup>190</sup>. Vi è però un altro risvolto di tale rigidità, che evidenzia il limite alla funzione preventiva, nel senso che non è da escludere che dal rispetto di quelle norme derivi una minaccia all'integrità dei beni giuridici. La funzione di prevenzione è allora capovolta: la struttura rigida da forma di sicura tutela si trasforma in occasione di offesa del bene giuridico<sup>191</sup>. Ciò significa che in certe situazioni è la stessa fattispecie concreta a richiedere di discostarsi dalla regola rigida, che non è in grado di neutralizzare gli effetti lesivi della condotta se non quando è violata: talvolta è proprio invadendo la corsia opposta che può essere evitato l'impatto con un pedone e il tamponamento con l'auto che si trova dietro e con la quale l'urto sarebbe inevitabile ove si eseguisse una brusca frenata per salvare il pedone. Si ritiene, allora, che la situazione concreta impone un uso "elastico" anche delle norme rigide proprio allo scopo di valorizzare appieno l'esigenza di una tutela effettiva della sicurezza individuale<sup>192</sup>.

Ma questa *valutazione diversa* della situazione di pericolo potrebbe fallire se lasciata alla macchina, portata ad arrestarsi anche bruscamente a costo di mettere in pericolo il

---

<sup>190</sup> G. Marinucci, *La colpa, cit.*, p. 247.

<sup>191</sup> *Ibidem*, p. 248.

<sup>192</sup> *Ibidem*, p. 244.

conducente della stessa auto e di quella che segue in coda, pur di arrestarsi dinanzi al percorso di un pedone maldestro.

Questo aspetto che qui analizziamo nel contesto della colpa penale e del dovere di diligenza si lega in realtà al problema dell'eticità dei sistemi di intelligenza artificiale su cui molto si dibatte<sup>193</sup>. La questione è resa molto concreta attraverso il noto dilemma del carrello ferroviario (*trolley problem*), che se originariamente si era posto come un dilemma per l'uomo alla guida del tram, con l'avvento della IA è stato presentato come dilemma per il veicolo stesso, chiamato ad assumere decisioni in situazioni di emergenza<sup>194</sup>.

L'esempio sollecita il dibattito intorno all'idea di macchine dotate di una coscienza e quindi anche di una propria etica, idea che però è messa in discussione a causa delle capacità ridotte delle tecnologie attualmente a disposizione, tra cui anche il deep learning, che offrono risultati sulla base di un calcolo statistico, individuando correlazioni in una mole di dati che un essere umano non saprebbe esaminare con la stessa velocità e precisione, ma che, a differenza degli uomini, non hanno contezza del significato delle loro scelte. Pertanto, il veicolo a guida autonoma che si trovi a dover scegliere tra due possibilità che comunque causeranno un danno a qualcuno, per ritornare all'esempio sopra proposto, si dice che non può essere considerato come un agente che si trovi di fronte a un dilemma morale<sup>195</sup>.

Certo, questo problema potrebbe essere superato se la scienza riuscirà a dotare la macchina di un codice etico, già impostato dal programmatore e inserito nel sistema secondo un modello *top-down*, oppure addestrando la macchina ad affrontare le scelte alla luce di valori etici, secondo un modello *bottom-up*. Ma anche a ritenere che ciò un giorno sarà possibile, ci si troverà dinanzi a sfide di altro tipo, come giustamente

---

<sup>193</sup> In argomento M. Anderson, S.L. Anderson (a cura di), *Machine Ethics*, Cambridge, Cambridge University Press, 2011; W. Wallach, C. Allen, *Moral Machines: Teaching Robots Right from Wrong*, Oxford, Oxford University Press, 2009; D. Tafani, *Sulla moralità artificiale. Le decisioni delle macchine tra etica e diritto*, in *Riv. di filosofia*, 1, aprile 2020, pp. 83 ss..

<sup>194</sup> Come ricorda D. Tafani, *cit.*, l'esempio fu presentato per la prima volta dalla filosofa inglese Philippa Ruth Foot, nel saggio sull'aborto dal titolo *The Problem of Abortion and the Doctrine of the double Effect*, pubblicato nella rivista *Oxford Review*, V, 1967, pp. 5-15 e poi ripreso e riproposto col nome di "trolley problem" da Judith J. Thomson, *Killing, letting die and the trolley problem*, in *The Monist*, LIX, 1976, pp. 204-217. Si immagina in sostanza il guidatore di un tram fuori controllo, che ha la sola possibilità di deviare il tram in un binario diverso da quello che sta percorrendo, ma ecco il dilemma: cinque uomini lavorano su un binario e solo un uomo sull'altro e chiunque si trovi sul binario in cui entrerà il tram è destinato a essere ucciso. La domanda allora è: uccidere cinque uomini che si trovano sulla traiettoria che è già percorsa dal tram o ucciderne uno che si trova sul binario verso il quale il tram può essere deviato?

<sup>195</sup> D. Tafani, *cit.*, p. 84.

avvertono i filosofi del diritto<sup>196</sup>. Anzitutto non esiste una sola etica, ma più etiche, al plurale, nelle quali sono trasposti i valori culturali, religiosi, deontologici che variano a seconda delle zone del mondo interessate. Ma non è finita qui: l'etica non è qualcosa di statico, ma varia nelle epoche storiche e quindi un codice etico dovrebbe essere sottoposto a un aggiornamento costante. Infine, l'etica è chiamata a risolvere i conflitti tra i valori: è possibile risolvere una volta per tutte questi conflitti per trasporli in una formula matematica<sup>197</sup>? Si comprende, allora, come il problema non sia solo quello di trovare una "formula matematica" etica, perché a monte la questione è quella di individuare un'etica per la matematica. Tra l'intervenire sui sistemi per renderli etici, da un canto, e l'individuazione dei doveri di chi li progetta e di chi ne fruisce, dall'altro, si condivide qui l'opinione di chi sostiene che la strada attualmente da percorrere con più urgenza è la seconda, ferma restando l'importanza di sostenere anche economicamente studi per un'etica dei robot<sup>198</sup>.

Torniamo allora al problema che ci eravamo posti: come rendere possibile la violazione necessaria della regola cautelare da parte di un sistema che si atterri pedissequamente al suo rispetto? L'unico modo mi pare possa essere la valorizzazione del ruolo dell'utente, che spiegherebbe e darebbe un senso a una visione di IA antropocentrica. L'agente umano potrà essere in grado di comprendere che la situazione

---

<sup>196</sup> S. Amato, *Biodiritto 4.0*, cit., pp. 103 ss.

<sup>197</sup> Si tratta infatti di decisioni spesso molto delicate, che possono dare spazio a pregiudizi anche discriminatori. Per tornare all'esempio del *trolley problem*, potremmo forse ritenere più condivisibile la scelta di deviare il tram sul secondo binario per uccidere solo una persona, anziché cinque, e approvare questa scelta argomentando anche alla luce delle caratteristiche fisiche di quell'uomo che è un uomo grasso e quindi meno utile alla società perché in sovrappeso? (il problema è evocato già nel titolo dal libro di D. Edmonds, *Uccideresti l'uomo grasso? Il dilemma etico del male minore*, Milano, Raffaello Cortina Editore, 2014). Per risolvere questo come altri dilemmi etici che si presentano nel caso delle auto di guida autonoma, è stata ideato un sito ([www.moralmachine.mit.edu](http://www.moralmachine.mit.edu)) che raccoglie le opinioni degli utenti messi di fronte a casi ipotetici come salvare una madre con un bambino all'interno dell'auto o un pedone. Nel sito lo scopo della raccolta dati è così descritto: "*This website aims to take the discussion further, by providing a platform for 1) building a crowd-sourced picture of human opinion on how machines should make decisions when faced with moral dilemmas, and 2) crowd-sourcing assembly and discussion of potential scenarios of moral consequence*". Ma come avverte chi ha riflettuto sul tema, la creazione di una "gerarchia di bersagli" dei veicoli a guida autonoma funzionale alla creazione di "algoritmi sacrificali" non trova risposta nel diritto perché la risposta alla domanda su chi uccidere "non si trova nel diritto, perché la domanda stessa è vietata", D. Tafani, cit., p. 89. L'alternativa a un sondaggio mondiale per la raccolta di soluzioni a dilemmi morali è quella di dotare i veicoli a guida autonoma di dispositivi etici personalizzati, una sorta di "manopola etica" che rimette al conducente la scelta, la quale è poi salvata e messa in atto in caso di incidenti G. Contissa, F. Lagioia, G. Sartor, *The Ethical Knob: Ethically customisable automated vehicles and the law*, in *Artificial Intelligence and Law*, XXV, 3, 2017, pp. 365-378; *La manopola etica: i veicoli autonomi eticamente personalizzabili e il diritto*, in *Sistemi intelligenti*, III, 2017, pp. 601-14.

<sup>198</sup> È questa l'opinione di S. Amato, cit., p. 107, il quale sottolinea che allo sviluppo di una dimensione morale delle macchine dovrebbero essere indirizzati i finanziamenti pubblici, a fronte di un mercato che già investe ampiamente su tutti gli altri aspetti del settore della IA.

concreta richiede la violazione della regola cautelare rigida; è in questi momenti allora che sarebbe richiesta la ripresa di controllo del sistema da parte dell'utilizzatore, chiamato ad adeguare il proprio comportamento alla più generale regola di diligenza che le peculiarità della fattispecie concreta rendono prevalente rispetto alla regola a struttura rigida. Per ritornare al nostro esempio, solo il conducente potrà ritenere opportuno invadere la corsia contrapposta per evitare lo scontro col pedone e l'impatto con l'auto posteriore.

Questo è il risultato imposto da macchine non pensanti, ma che calcolano. In questi casi, la lesione provocata al conducente della vettura che seguiva quella dotata di pilota automatico, con la quale avviene il tamponamento a causa del brusco arresto che il sistema effettua per evitare lo scontro con il pedone, non potrà a mio avviso essere imputata al produttore, che ha fornito il manufatto intelligente di tutte le conoscenze disponibili *a priori*. Potrà invece essere valutata la colpa del conducente della vettura con pilota automatico, al quale potrà essere contestato di non aver supervisionato le azioni del sistema di guida autonoma e di non aver evitato l'evento ponendo in essere una manovra di invasione della corsia opposta.

## **8. Diligenze a confronto**

Si è cercato di riflettere sulle capacità di un sistema intelligente e la conclusione raggiunta è che il comportamento diligente del sistema artificiale non potrà mai essere equiparato alla diligenza umana.

La diligenza artificiale è limitata al campo di conoscenza di cui dispone il sistema, il quale non sarà dotato di spirito critico, di buonsenso, di capacità di valutazione complessiva della situazione di pericolo, che consentono la concretizzazione delle regole cautelari, specialmente quelle elastiche e generiche, nel caso concreto. La diligenza artificiale esiste con riguardo a un campo limitato, cioè entro quell'ambiente conoscitivo che è costruito dal produttore. All'interno di quello spazio di conoscenza, il sistema artificiale saprà calcolare la soluzione migliore cui conformare il proprio comportamento, ma non appena la situazione di pericolo fuoriesce dal perimetro della conoscenza artificiale, la macchina non avrà risposte e dovrà procedere generalizzando a partire dagli esempi conosciuti e dai dati in suo possesso, col rischio di un errore.

Naturalmente, i limiti della diligenza artificiale rilevano ai fini della valutazione della condotta colposa del produttore e dell'utilizzatore (e solo a tali fini, almeno fintantoché sarà esclusa la responsabilità singolare dei sistemi di IA).

Avevamo sopra distinto le situazioni di pericolo connesse al mancato rispetto di regole cautelari in situazioni di cui è possibile fornire *ex ante* una descrizione in termini di pericolosità e situazioni in cui tale descrizione *ex ante* della pericolosità risulta impossibile, poiché correlate a dinamiche concrete difficilmente prevedibili e quindi non immaginate nella fase di addestramento dell'algoritmo e proposte come esempi del *training set*. In altri termini, si tratta di casi che non possono essere codificati nel linguaggio computazionale per renderli conoscibili in anticipo alla macchina, perché non preventivabili. In presenza di tali circostanze (che rientrano nelle situazioni-tipo sub c) dello schema proposto *supra*) il sistema procederà con le generalizzazioni che possono contenere un margine di errore che non era evitabile dal produttore nemmeno adoperando un addestramento diverso, data l'imprevedibilità della dinamica concreta.

Ne deriva che il produttore non può essere chiamato a rispondere per una generalizzazione che non sia adatta per il caso concreto se -e solo se- viene in gioco un rischio del tutto imprevedibile, originato dalla combinazione di una serie di fattori verificatisi nel caso concreto, che rendono l'intera dinamica che porta alla causazione dell'evento assolutamente peculiare e unica.

Le situazioni non codificabili in linguaggio informatico *ex ante* dal produttore potranno però, almeno in alcuni casi, essere riconoscibili e evitabili dall'utente, chiamato a vigilare sul corretto funzionamento del sistema e a intervenire ove necessario. A questo punto sarà necessario l'intervento dell'agente umano che dispone di una capacità di lettura semantica, è cioè in grado di comprendere la situazione di pericolo nella sua concretezza, con le peculiarità che essa presenta e per questo potrà adoperare una condotta che sia la più diligente non in astratto, ma in concreto.

Tornando al caso dell'auto a guida autonoma che attraversando col verde si scontra con l'auto proveniente lateralmente il cui conducente non si è arrestato a causa di un guasto che non ha fatto scattare il semaforo rosso, potremmo immaginare che al produttore del sistema di guida autonoma sarà contestato il non aver dotato l'auto di sensori migliori (sempreché disponibili nel mercato, secondo la logica tipica di un metodo deontologico dell'accertamento del parametro obiettivo di diligenza) in grado di percepire anche i movimenti improvvisi, quindi una colpa potrà sempre essere riconosciuta, ma dovrà essere attenuata in ragione dell'imprevedibilità del rischio e del preventivo rispetto

da parte del sistema della regola cautelare che afferma che è possibile proseguire la marcia in presenza del segnale luminoso verde.

Orbene, è vero che queste stesse argomentazioni possono valere per escludere la responsabilità dell'utente, ma come si è cercato di dimostrare, la loro portata è diversa. L'individualizzazione del giudizio di colpa in sede di misurazione dello standard di diligenza esigibile porta a tener conto delle maggiori conoscenze causali dell'agente le quali, se presenti, comportano un dovere di intervento dell'utente sul controllo del sistema, non potendosi più accettare una sua condotta diligente limitata al dovere di vigilanza.

Al produttore e, più in particolare, al programmatore e all'addestratore dell'algoritmo, non potrà essere mosso un rimprovero per non aver dotato il sistema di conoscenze sufficienti a comprendere appieno il caso concreto, le cui peculiarità resteranno incomprensibili al sistema il quale è per sua natura dotato di una conoscenza limitata. La difettosità del prodotto non può essere fondata sull'assenza di istruzioni su come reagire in presenza di situazioni del tutto peculiari perché ciò significherebbe richiedere una prestazione irrealizzabile al produttore, consistente nel prevedere tutto ciò che può accadere nel corso dell'utilizzo del sistema di IA e in qualsivoglia contesto esso venga impiegato (per rimanere nell'ambito della guida autonoma, si dovrebbe distinguere a seconda dell'ambiente in cui la vettura è utilizzata, per prevedere quali possano essere i rischi collegati alla presenza di pedoni, di animali selvatici e di che tipo di animali ecc...).

A ciò si aggiunga che regole generali di diligenza, che consentirebbero a un agente umano immaginato al posto della macchina di riconoscere la situazione di pericolo, sono difficili da codificare in un linguaggio computazionale e spesso non possono avere alcuna utilità se inserite nel sistema di IA, che non sarebbe comunque in grado di declinarle nel caso concreto. Per questo, anche ove non inserite, non penso si possa muovere un rimprovero al produttore per aver costruito un prodotto difettoso.

I limiti dell'intelligenza artificiale qui descritti servono anche per escludere a monte lo scenario di una responsabilità singolare del sistema di IA. Le peculiarità che caratterizzano questi nuovi prodotti, soprattutto la capacità di autoapprendimento, devono sicuramente incentivare la riflessione su un adeguamento delle norme esistenti, ma non rendono necessaria, almeno per ora, la creazione di nuovi centri di imputazione giuridica, sembrando sufficiente uno sforzo che si concentri a regolare le condotte delle persone che producono e utilizzano tali sistemi e interagiscono con essi, prestando attenzione alla diversificazione delle responsabilità (ad esempio nella filiera produttiva), nell'idea, condivisa qui, che non sia allo stato attuale necessario un diritto per i robot, ma una

regolamentazione della robotica<sup>199</sup>, alla cui elaborazione anche il diritto penale è chiamato a contribuire.

---

<sup>199</sup> S. Amato, *op. cit.*, p. 91.



## SEZIONE II

### LA CAUSALITÀ

SOMMARIO: 1. “Non una semplice coincidenza, ma una vera relazione di causalità”. Il criterio nomologico di accertamento della causalità come espressione del principio di legalità- 2. Le leggi scientifiche - 3. La causalità generale e l’aumento del rischio - 3.1 La probabilità nelle tecnologie di IA: margine di errore e *accuracy* come strumenti a servizio della decisione giudiziaria - 3.2 Indagine epidemiologica e risultati del test di validazione dei sistemi di IA: (tante) analogie e (poche) differenze - 4. La causalità individuale o specifica - 4.1 L’incertezza: la causalità individuale rinchiusa in una scatola nera. Opacità *ex post* e *explainable IA* - 4.2 Verso il superamento dell’opacità causale? - 5. Brevi riflessioni sulla causalità omissiva - 5.1 Alcuni aspetti problematici. La difficile distinzione tra azione e omissione - 5.2 (segue) La prova per esclusione – 6. Nuove prospettive

#### **1. “Non una semplice coincidenza, ma una vera relazione di causalità”.**

##### **Il criterio nomologico di accertamento della causalità come espressione del principio di legalità**

Nell’analisi del rapporto tra lesione o offesa di beni giuridici derivanti da produzione o uso dei sistemi di IA e responsabilità penale, va investigato anche il nesso di causalità, in quanto elemento coesistente di un diritto penale a base oggettivista, contrapposto al diritto penale soggettivistico, in cui il problema eziologico si sovrappone e si confonde con quello della colpevolezza o della pericolosità del reo<sup>1</sup>.

Del resto, è la stessa Costituzione, all’art. 27, a imporre un’investigazione sull’attribuibilità del fatto criminoso al soggetto allorché sancisce che “la responsabilità penale è personale”, attribuibilità che va verificata prima sul piano oggettivo e solo dopo su quello soggettivo affinché l’evento possa essere ascritto all’agente come fatto suo proprio<sup>2</sup>. Sovviene, a tale scopo, la causalità, intesa come aspetto della tipicità che studia il legame tra l’azione o l’omissione e l’evento, che rappresenta il “requisito minimo indispensabile” sul piano della fattispecie obiettiva<sup>3</sup>.

Il problema della causalità si è storicamente presentato quando i giuristi hanno iniziato a porre l’attenzione non più solo sull’azione, ma anche sul risultato esteriore di essa, cioè l’evento, riconoscendo che quest’ultimo deve essere opera dell’uomo al pari dell’azione<sup>4</sup>.

---

<sup>1</sup> F. Mantovani, *Diritto penale, op. cit.*, p. 136.

<sup>2</sup> D. Pulitanò, *Diritto Penale, VIII ed.*, Torino, Giappichelli, 2019, p. 163.

<sup>3</sup> *Idem*. Nello stesso senso, F. Stella, *Leggi scientifiche e spiegazione causale nel diritto penale*, Milano, Giuffrè, ristampa, p. 101.

<sup>4</sup> F. Antolisei, *Il rapporto di causalità nel diritto penale*, Padova, CEDAM, 1934, p. 3. Poiché la causalità è un nesso che dall’azione si dirige verso l’evento, il problema del rapporto di causalità non ha importanza per tutto il campo del diritto e non è un requisito essenziale di tutti i reati, ma solo di quelli nei quali un

La difficoltà di determinare la natura del legame che deve sussistere affinché un evento possa oggettivamente imputarsi all'agente ha posto la questione del rapporto di causalità, che è resa ancor più complessa quando l'evento è il risultato di una pluralità di condizioni, le quali, pertanto, non si esauriscono nella condotta umana.

La questione che in questa sede si tenta di analizzare ci chiede di riflettere sul contributo che al decorso eziologico può essere apportato da forze *artificiali*, diverse da quelle umane, naturali o animali, e riconducibili a un sistema di IA. Pertanto, la questione che qui interessa è: cosa accade se tra l'azione o l'omissione dell'uomo e l'evento si inserisce un sistema di IA? È questo contributo rilevante ai fini dell'accertamento del nesso di condizionamento? E ancora: cosa accade se si accerta la rilevanza causale del sistema di IA, ma non si riesce a rintracciare una condotta umana causalmente rilevante?

Si pensi a casi, già noti alle cronache, di incidenti stradali che involgono auto a guida autonoma: l'auto a guida autonoma non riconosce un ostacolo, il rimorchio di un camion, e vi si schianta contro provocando la morte del conducente<sup>5</sup>; la *driverless car* non riconosce un pedone che attraversa le strisce pedonali, il quale muore a seguito dell'incidente<sup>6</sup>; Tizio e Caio decidono di affidarsi totalmente al sistema di guida autonoma e nessuno di loro siede al posto del conducente, l'auto esce di strada mentre affronta una curva e entrambi i passeggeri muoiono<sup>7</sup>.

E gli esempi potrebbero continuare.

La produzione e l'utilizzo di sistemi intelligenti sollevano non poche questioni che involgono la dimostrazione dell'enunciato causale "x ha causato y". È noto che tale dimostrazione è già in sé problematica, a prescindere dal contributo che possa dare un

---

evento è richiesto per la consumazione, ovvero sia condizione di punibilità del reato o circostanza aggravatrice della pena, Id. pp. 8-10.

<sup>5</sup> Fatto accaduto in Florida, nel 2016, come riportato da *ansa.it*, Primo incidente mortale alla guida di un'auto autonoma, 1 luglio 2016: "La Tesla si è andata a schiantare contro un tir proveniente dalla direzione opposta, che stava svoltando a sinistra. Tesla ha spiegato sul suo sito che né il conducente né il sistema di guida autonoma hanno visto il lato bianco del tir, che si trovava perpendicolare all'auto, contro un cielo particolarmente luminoso. "La particolare altezza del rimorchio - ha scritto l'azienda - combinata con la sua posizione attraverso la strada e le circostanze estremamente rare dell'impatto, hanno fatto sì che il modello S passasse sotto il rimorchio".

<sup>6</sup> Fatto accaduto in Arizona, quando la Volvo XC 90 con sistema *driverless* di Uber ha investito una donna che, secondo le ricostruzioni delle autorità, avrebbe attraversato di colpo, al di fuori delle strisce e portando a mano la sua bici. *Uber, incidente mortale per l'auto a guida autonoma*, *www.repubblica.it*, 20 marzo 2018

<sup>7</sup> *Tesla e guida autonoma, nuovo incidente: «Non c'era nessuno al volante»*, in *www.motori.corriere.it*, 19 aprile 2021, fatto accaduto a Nord di Houston, in Texas.

sistema di IA, in quanto mentre  $x$  e  $y$ , cioè la causa e l'effetto, sono accadimenti empirici percepibili, verificabili e riproducibili, tale non è il nesso causale che connette  $x$  e  $y$ <sup>8</sup>.

Ai fini della determinazione del nesso in questione, ricorriamo alla teoria condizionalistica, che nel nostro codice è accolta agli artt. 40 e 41 c.p., anche detta della *condicio sine qua non* oppure teoria dell'equivalenza delle condizioni, poiché considera equivalenti le concause che appaiano condizioni necessarie ma non necessariamente sufficienti<sup>9</sup>.

Tra le critiche mosse a questa teoria vi è quella dell'efficacia euristica limitata<sup>10</sup>, che mette in luce un problema di *prova* della causalità che, nell'ambito del diritto penale sostanziale, viene studiato con riferimento ai criteri pertinenti alla determinazione di cosa sia causale e cosa no<sup>11</sup>. E, in effetti, la formula della condizione *sine qua non* è stata ritenuta una “formula vuota” poiché nulla dice in ordine al criterio da utilizzare per stabilire il nesso di causalità, atteso che affermare che una certa azione è causa dell'evento se, eliminandola mentalmente, l'evento lesivo non si sarebbe verificato, “significa rifugiarsi dietro una vistosa tautologia”, mentre ciò che è necessario è conoscere il *criterio* per sostenere che, in assenza di quella condotta, l'evento non si sarebbe verificato<sup>12</sup>. Si deve a Engisch l'enunciazione di questo vizio di fondo della formula della *condicio* e il superamento attraverso il modello della condizione conforme a leggi, entro le quali l'evento va sussunto<sup>13</sup>, mentre in Italia il modello della sussunzione sotto leggi scientifiche, quale correttivo all'efficacia euristica limitata della regola della *condicio*, è stato recepito dalla dottrina e dalla giurisprudenza grazie all'opera di Federico Stella.

---

<sup>8</sup> M. Taruffo, *La prova del nesso causale*, in C. De Maglie, S. Seminara (a cura di), *Scienza e causalità*, Padova, CEDAM, 2006, p. 90 ss.

<sup>9</sup> D. Pulitanò, *Diritto Penale*, *op. cit.*, p. 164. Molti ordinamenti non contengono una disposizione generale sulla causalità, a differenza che nel nostro, dove è da questa base legislativa che bisogna prendere le mosse per lo studio del nesso causale, v. F. Antolisei, *Il rapporto di causalità*, *cit.*, pp. 145 ss. in cui si descrive la genesi del testo degli artt. 40 e 41 c.p. e la tesi che riconduce alle norme in oggetto la teoria della causalità adeguata, oltre che la tesi che sostiene che sia la base normativa della teoria condizionalistica.

<sup>10</sup> Accanto a questa, sono note le obiezioni del regresso all'infinito, della causalità alternativa ipotetica e della causalità addizionale, superabili, la prima col correttivo del dolo e della colpa, le altre due col correttivo del riferimento all'evento concreto, secondo la soluzione proposta in dottrina. Cfr. G. Fiandaca, E. Musco, *Diritto Penale, P.te generale*, VIII ed., pp. 245 – 247.

<sup>11</sup> D. Pulitanò, *Diritto Penale*, *op. cit.*, p. 165.

<sup>12</sup> A. Crespi, *Riforme ardite e riforme ordite*, *Riv.it. dir. proc. pen.*, 1973, pp. 819 e 820. Dello stesso tenore erano già le osservazioni di Grispigni sul progetto preliminare, riportate dall'Antolisei, *op. cit.*, p. 154, riferite al primo c. dell'art. 40, che “lascia insoluto il problema, che consiste nello stabilire quando è che un evento debbesi considerare conseguenza di un'azione, e cioè se basti aver posto una semplice condizione o sia necessaria una condizione, diremo così, qualificata”.

<sup>13</sup> K. Engisch, *Die Kausalität als Merkmal der strafrechtlichen Tatbestände*, Tübingen, Mohr, 1931, p. 13 ss.; M. Romano *Nesso causale e leggi scientifiche in diritto penale*, in *Scritti per Federico Stella*, Napoli, Jovene, 2007, p. 895.

Come osservava Engisch, il giudizio controfattuale che il giudice è chiamato a effettuare attraverso il procedimento di eliminazione mentale non dà informazioni sul collegamento tra l'azione e l'evento; la previa conoscenza del fatto che la condotta appartiene al novero delle azioni causa dell'evento presuppone che si conosca in anticipo una legge causale espressiva della regolarità tra fenomeni<sup>14</sup>. Così, se il nesso causale non può essere provato in modo diretto poiché non è un avvenimento empirico, esso può essere provato in via "inferenziale", cioè dimostrando l'esistenza di una legge di natura riferibile al caso di specie<sup>15</sup>.

Il richiamo alle "leggi di natura" che spieghino i legami eziologici determina che il sapere scientifico stia alla base del giudizio sulla causalità: ci si avvale di un metodo definito scientifico, poiché il caso concreto deve essere sussumibile sotto il sapere che deriva dalle leggi scientifiche<sup>16</sup> esplicative dei legami tra azione ed evento con regolarità stabilite dalla scienza e non dal giudice, il quale non può che essere un "consumatore" di leggi causali<sup>17</sup>. Vi è una connessione logica tra l'aver una conoscenza delle leggi e l'essere capace di dare spiegazioni che siano giustificabili<sup>18</sup>.

Così intesa, la formula della *condicio sine qua non* implica la nozione "nomologica" di causa<sup>19</sup>, poiché sottintende il rinvio al concetto di spiegazione mediante leggi ed

---

<sup>14</sup> M. Maiwald, *Causalità e diritto penale*, Milano, Giuffrè, 1999, p. 7; F. Mantovani, *Diritto penale*, *op. cit.*, p. 139. La ricerca della legge che spiegasse il legame causale tra azione ed evento è stata condotta nel processo del talidomide decisa dal Landgericht di Aquisgrana, come ricorda Maiwald, il quale si mostra sorpreso dal fatto che la dottrina non abbia riconosciuto prima tale necessità, atteso che il modo di procedere adottato in quel processo, consistente nel ricorso alla perizia per stabilire la sussistenza del nesso di causalità, appartiene alla prassi quotidiana delle decisioni dei tribunali, p. 8.

<sup>15</sup> M. Taruffo, *op. cit.*, p. 92.

<sup>16</sup> Anche dette 'leggi di natura', 'legge naturale' o semplicemente 'legge'. E. Nagel, E. Nagel, *La struttura della scienza*, Milano, Feltrinelli, 1968, p. 54. Nota l'A. che la terminologia non è tecnica e che non vi è una definizione univoca di cosa sia effettivamente una 'legge e quindi è inutile il tentativo di fornire una definizione univoca con grande precisione logica (pp. 55 -56), per poi giungere ad affermare che "in effetti le leggi di natura vengono comunemente usate per giustificare condizionali congiuntivi e controfattuali" (è un condizionale congiuntivo l'affermazione "per ogni  $x$ , se  $x$  è un corvo, allora  $x$  è nero). (p. 57).

<sup>17</sup> F. Stella, *Leggi scientifiche e spiegazione causale nel diritto penale*, Milano, Giuffrè, 2000 ristampa, p. 102. L'espressione è mutuata da C.B. Joynt e N. Rescher, *The problem of Uniqueness in History*, in *History and Theory*, I, 1961, n. 2, p. 154, i quali riferiscono l'idea di "consumatore di leggi" allo storico. Allo stesso modo W. H. Dray, *Filosofia e conoscenza storica*, Bologna, Il Mulino, 1973, p. 18, si riferisce allo storico come consumatore di leggi, evidenziando che nella storia "lo scopo della ricerca non è lo stabilire delle leggi", per questo "le leggi necessarie per la spiegazione dovrebbero derivare da altri campi", come le scienze sociali o dal senso comune. È allora Stella che riferisce l'espressione di "consumatore di leggi" al giudice. In quest'ultimo senso, *ex multis*, Cassazione penale sez. IV, 10/06/2010, n. 38991, secondo cui il giudice non solo non crea la legge, ma la rileva, ma deve anche "motivare la scelta della legge scientifica che ritiene più convincente ed idonea o meno a spiegare l'efficacia causale di una determinata condotta" (massima riportata in *Responsabilità Civile e Previdenza 2011*, 2, 346, nota di Coggiola).

<sup>18</sup> W. H. Dray, *op. cit.*, p. 18.

<sup>19</sup> F. Stella, *op. cit.*, p. 101. Per chiarezza espositiva impieghiamo l'espressione "nozione nomologica" e non "nomologico-funzionale" di causa, pure utilizzata da Stella, poiché nella letteratura scientifica la cd.

esclude che la determinazione del nesso di condizionamento possa essere affidata al libero apprezzamento soggettivo del giudice<sup>20</sup>. Qualcuno commenta questa rilettura della formula della *condicio* osservando che essa è ormai del tutto sostituita dalla spiegazione nomologico-scientifica, col risultato che la prima resta un giudizio fondato su un mero ragionamento logico, che fornisce una spiegazione *ex ante* e solo apparentemente *ex post*, perché il giudizio di eliminazione mentale viene effettuato sapendo già il valore causale di esso sulla base delle leggi scientifiche, per cui la formula della *condicio* non aggiungerebbe nulla alla ricostruzione della catena causale<sup>21</sup>.

Al di là di tali osservazioni, il vero merito che va riconosciuto al rinvio alle leggi di natura è quello di aver conformato l'accertamento del nesso di causalità al rispetto di due principi fondamentali del diritto penale, contemplati dalla Carta costituzionale: il principio di legalità in materia penale e quello di personalità della responsabilità penale. Se infatti si concorda nel ritenere che non esiste una nozione unica di causa, ma che essa varia a seconda del *contesto* in cui viene in rilievo<sup>22</sup>, è doveroso leggere l'accezione penalmente rilevante di causa in modo coerente con i fondamentali principi di garanzia del diritto penale moderno.

Il rinvio a leggi di natura richiama il principio di legalità poiché, come notava Crespi, è solo in questo modo che possono essere fissati “una volta per tutte” i criteri di determinazione della causalità che vincolano il giudice escludendo sue valutazioni discrezionali, così da ritenere rispettati i principi di legalità e di certezza del diritto anche in tema di causalità<sup>23</sup>. La posta in gioco è alta: si tratta della funzione di garanzia della fattispecie<sup>24</sup>, di cui il nesso di causalità è un requisito che, in quanto tale, non può essere configurato in modo da rimanere indeterminato<sup>25</sup>.

---

“nozione nomologico-funzionale” è riferita all'impiego di sole leggi universali o ineccepibili (ad esempio questo è l'utilizzo che ne fa A. Pasquinelli, *Nuovi principi di epistemologia*, Milano, Feltrinelli, 1974, a p. 101), come si preoccupa di precisare lo stesso Stella (*op. cit.* p. 97, nota 48), chiarendo subito che nel diritto penale il “perché” dell'evento non può essere individuato solo mediante il riferimento a leggi ineccepibili o universali (p. 98, nota 49). L'espressione “nozione nomologica di causa” è quindi più generica, implicando il riferimento a delle leggi scientifiche, senza escludere quelle statistiche. Il paradigma della sussunzione sotto leggi scientifiche è stato recepito dalla giurisprudenza a partire dalla sentenza della Corte di Cassazione sul disastro di Stava del 1985, (Cass., 6 dic. 1990, *Foro it.*, 1991, II, pp. 62 ss.), come ricorda D. Pulitanò, *Gestione del rischio da esposizioni professionali*, Cass. pen. 2, 2006, p. 789.

<sup>20</sup> F. Stella, *Leggi scientifiche*, *op. cit.*, p. 101.

<sup>21</sup> R. Bartoli, *Il problema della causalità penale*, Torino, Giappichelli, 2010, p. 55.

<sup>22</sup> F. Stella, *Il giudice corpuscoliano. La cultura delle prove*, Milano, Giuffrè, 2005.

<sup>23</sup> A. Crespi, *op. cit.*, p. 820.

<sup>24</sup> *Idem*, p. 820.

<sup>25</sup> F. Stella, *Leggi scientifiche*, *op. cit.*, p. 90.

E infatti, come osserva Stella, vincolando tale accertamento a una legge di copertura che spieghi perché un dato evento è riconducibile a quella determinata condotta, la formula della *condicio sine qua non* viene interpretata in modo conforme al principio di tassatività della fattispecie (e solo questa e nessun'altra interpretazione potrebbe conformare l'art. 40 c.p. al rispetto di tale principio)<sup>26</sup>, in grado di rendere il nesso di condizionamento un requisito non indeterminato e indeterminabile, ma, al contrario, determinato, poiché il suo contenuto è stabilito di volta in volta attraverso l'indicazione delle condizioni empiriche antecedenti in grado di produrre l'evento, le quali costituiscono la causa nomologicamente intesa<sup>27</sup>. Va però chiarito che la rilevanza delle leggi scientifiche ai fini di una piena attuazione del principio di legalità non comporta che le prime diventino elemento costitutivo della fattispecie<sup>28</sup>. Elemento costitutivo è e resta il nesso causale che è elemento astratto della fattispecie legale, il quale viene spiegato in concreto attraverso il richiamo alle leggi causali. Per questa ragione le leggi causali si sottraggono alle regole sulla successione di leggi penali nel tempo, dovendo il giudice pacificamente stabilire il nesso causale sulla base delle conoscenze scientifiche presenti al momento del giudizio e non del fatto<sup>29</sup>.

Inoltre, la regolarità causale condensata nella legge scientifica serve a rendere chiaro il confine tra ciò che è lecito e ciò che non lo è. In questo senso il principio di legalità viene in rilievo come prevedibilità del risultato, cioè della sanzione, con l'effetto di valorizzare la legalità tanto nella sua declinazione in principio di tassatività, come osservava Stella, quanto nella sua dimensione europea, che considera il principio di legalità come pre-calcolabilità dell'intervento punitivo, come più volte ribadito dalla Corte di Strasburgo, la quale legge il '*nullum crimen, nulla poena sine lege*' come diritto dell'individuo, anziché come principio<sup>30</sup>. Il principio di prevedibilità della risposta

---

<sup>26</sup> F. Stella, *op. cit.*, p. 102.

<sup>27</sup> *Idem*, p. 101.

<sup>28</sup> M. Romano, *Nesso causale, cit.*, p. 898.

<sup>29</sup> M. Donini, *Il garantismo della condicio sine qua non, cit.*, p. 497 e 529; P. Veneziani, *Il nesso tra omissione ed evento nel settore medico: struttura sostanziale ed accertamento processuale*, in *Studi in Onore di Giorgio Marinucci*, Milano, Giuffrè, 2006, p. 1981.

<sup>30</sup> F. Viganò, *Il principio di prevedibilità della decisione giudiziale in materia penale*, *Dir. pen. cont.*, 19 dicembre 2016, p. 12. Va qui ricordato che la prevedibilità, assieme all'accessibilità, è un concetto utilizzato per fare riferimento alla legge intesa non solo come dato testuale, ma anche come "norma vivente" resa tale dall'interpretazione giurisprudenziale, sul punto si rinvia a G. De Amicis, *Il principio di legalità nella giurisprudenza delle Corti europee*, in *Quaderni europei*, ottobre 2009, n. 14, p. 6; G. Grasso, F. Giuffrida, *Gli effetti della giurisprudenza della Corte Edu sull'ordinamento italiano: prospettive di diritto penale sostanziale*, *L'indice penale*, Gennaio-aprile 2020, p. 52.

punitiva è invocato in questa sede in quanto va riferito alla norma penale in tutte le sue componenti, quindi anche nella sua dimensione causale. E infatti, la sanzione penale non sarà certamente prevedibile se rinunciamo all'impiego di una legge di copertura per rimettere l'accertamento del primo momento della responsabilità penale, cioè il momento causale, all'intuito del giudice.

La prevedibilità della risposta sanzionatoria, garantita da un fatto tipico criminoso che sia chiaro anche nella sua componente causale, arricchisce dipoi la norma penale del compito di guidare i comportamenti dei consociati i quali, se decideranno di contravvenire al divieto penale, lo faranno essendo consapevoli delle conseguenze cui andranno incontro tenendo quella determinata condotta. In assenza di tale consapevolezza, la pena irrogata a un soggetto al quale non è possibile muovere alcun rimprovero perderebbe la funzione rieducativa assegnatale dal comma 3 dell'art. 27 Cost. Viene qui in rilievo il principio della responsabilità penale personale, inteso non solo come divieto di responsabilità per fatto altrui, ma anche come rimproverabilità del soggetto per aver tenuto un comportamento colpevole<sup>31</sup>.

Il principio di personalità della responsabilità penale reclama una valutazione, già sul piano causale, che sia la più certa possibile e questa esigenza è soddisfatta dal sapere scientifico esplicativo, in quanto sapere che offre il massimo grado di certezza<sup>32</sup>. Se la responsabilità è per fatto "proprio" nel momento in cui l'evento è da quel soggetto cagionato, allora l'accertamento causale deve essere rigoroso e il ricorso a leggi scientifiche, attraverso il richiamo a regolarità scientifiche che governano gli accadimenti, soddisferebbe l'esigenza di un riscontro effettivo della connessione causale rispetto all'evento<sup>33</sup>.

Ma, a ben vedere, un accertamento causale sganciato da una base scientifica ha conseguenze anche sociali: finché la decisione risponde a istanze general-preventive (o anche punitive) condivise in una comunità noncurante del ruolo svolto dal principio di legalità e di responsabilità personale in materia penale, nulla sarà obiettato dall'opinione pubblica (ma non certo dagli studiosi) al giudice che condanni l'imputato pur in assenza

---

<sup>31</sup> G. Fiandaca, E. Musco, *Diritto penale, Parte generale, cit.*, pp. 338 ss.

<sup>32</sup> R. Bartoli, *Diritto penale e prova scientifica*, in G. Canzio, L. Luparia (a cura di), *Prova scientifica e processo penale*, Padova, CEDAM, 2018, p. 86

<sup>33</sup> Sulla rilevanza dell'accertamento causale per l'affermazione del principio di responsabilità personale per fatto proprio, si veda M. Donini, *La causalità omissiva e l'imputazione "per l'aumento del rischio"*, in *Riv. it. dir. proc. pen.*, 1999, p. 39, il quale critica il ricorso a parametri alternativi a quelli eziologici, in particolare i criteri basati sul mero rischio.

di un robusto addentellato scientifico che dimostri il legame causale tra la condotta e l'evento. Le reazioni saranno però diverse nel momento in cui la decisione giudiziale risulti incomprensibile all'opinione pubblica, che si schiera contro la decisione finale, sia essa di condanna o di assoluzione. In casi simili, in cui viene a mancare la base emotiva della collettività a sostegno della sentenza, solo un'argomentazione scientificamente fondata in punto di causalità potrà risultare plausibile e condivisibile, quantomeno sul piano del diritto, e quindi potrà essere accettata da chi vorrà comprenderne le ragioni. Quelle ragioni, infatti, saranno spiegabili in modo scientifico e rigoroso e ciò consente all'opinione pubblica una valutazione autonoma sul potere di convincimento degli argomenti proposti e il controllo della decisione da parte degli organi giudiziari superiori<sup>34</sup>.

La conclusione è allora obbligata nel senso di escludere il ricorso al metodo delle scienze spirituali, che contrappongono alla regolarità causale condensata nella legge scientifica il convincimento soggettivo del giudice<sup>35</sup>. Questo metodo troverebbe fondamento sul terreno della "logica giudiziale" ed è stato invocato dalla giurisprudenza tedesca nel noto caso sulle malformazioni da talidomide<sup>36</sup>. In quella circostanza, i giudici tedeschi, rilevato che le risposte della scienza erano insufficienti, giunsero a sostenere che la prova delle scienze naturali, ove non presente, è superflua nel contesto processuale, dove il legame eziologico può ritenersi provato col metodo delle scienze spirituali che consente al giudice di approdare a una certezza soggettiva. Il risultato è la negazione della rilevanza del principio di legalità in sede di accertamento del nesso eziologico. Fare riferimento all' 'intuizione' o al 'fiuto' del giudice, interpretando la formula della *condicio sine qua non* nel suo significato "individualizzante", significa abbandonare criteri generali vincolanti per il giudice nella determinazione di un requisito della fattispecie, per lasciare spazio alle sue libere valutazioni soggettive<sup>37</sup>. Ragionando in questo modo, i giudici disattenderebbero "impunemente" il principio di legalità, "arrogandosi il diritto di determinare liberamente, e di volta in volta, taluni requisiti del fatto tipico": questo il commento duro di Stella sulle sentenze sul disastro del Vajont<sup>38</sup>,

---

<sup>34</sup> M. Maiwald, *op. cit.*, pp. 111-112.

<sup>35</sup> F. Stella, *op. cit.*, p. 78; G. Licci, *op. cit.*, pp. 146 ss.

<sup>36</sup> Caso affrontato dalla giurisprudenza tedesca, ampiamente studiato e approfondito da Stella, *Leggi scientifiche, cit.*, pp. 33 ss., commentato da H. Sjöström, R. Nilsson, *Il talidomide e il potere dell'industria farmaceutica*, Milano, Feltrinelli, 1973.

<sup>37</sup> F. Stella, *op. cit.*, pp. 73-76

<sup>38</sup> F. Stella, *Leggi scientifiche, op. cit.*, p. 77. Si tratta della sent. Corte d'App. Aquila, 3 ottobre 1970, *Rass. Giur.* 1971 p. 101, e Cass. 25 marzo 1971, *ENEL, Rass. Giur.* 1971, p. 649.

dove si rinunciò alla prova della relazione causale tra la frana e l'opera dell'uomo sulla base di leggi di natura.

Sempre per effetto del necessario fondamento nomologico che deve avere l'accertamento del nesso causale, non è condivisibile una prova di tale legame eziologico basata sullo stesso accadimento dei fatti, cioè sulla loro successione cronologica nel caso di specie. Questo è quanto sembra sia stato fatto nel caso delle "macchie blu", anche se, a dispetto di espressioni linguistiche chiare nel senso di rintracciare la causalità sulla base del solo rapporto tra i fatti, a ben vedere i giudici tentarono di stabilire se vi fosse o meno l'esistenza di una regolarità tra l'inalazione dei fumi della fabbrica di alluminio e l'insorgere di manifestazione morbosa cutanea<sup>39</sup>. Un ragionamento causale basato sull'idea per cui "*post hoc, propter hoc*" non è accoglibile, poiché il giudice penale non deve andare alla ricerca di "una semplice coincidenza", ma è sempre necessario appurare "una vera relazione di causalità"<sup>40</sup>.

Infine, è da ripudiare pure il valore esplicativo delle *massime di esperienza*<sup>41</sup>, da intendersi come le generalizzazioni della vita quotidiana o il senso comune. In effetti, il senso comune si contrappone alla funzione di garanzia della fattispecie e arriva a negarla se si considera che il senso comune non sempre è empiricamente verificabile. Va però tenuto in considerazione che talvolta il senso comune può presupporre una legge scientifica: in tale evenienza esso si limiterebbe a semplificare l'espressione linguistica dell'enunciato della legge causale<sup>42</sup>. Il risultato è che la spiegazione causale ha solo la forma della massima di esperienza, mentre presuppone una legge scientifica e, pertanto,

---

<sup>39</sup> È quanto osserva Stella, laddove afferma che: "l'ambiguità del (nostro) linguaggio è spesso causa di numerosi fraintendimenti; e proprio in essa va probabilmente ricercata la ragione per la quale, nel caso in esame, non ci si è resi completamente conto che, nel prendere in considerazione e nel raffrontare fra loro i fatti suindicati, non si faceva altro che cercare di stabilire l'esistenza o meno di una regolarità", F. Stella, *Leggi scientifiche, op. cit.* p. 50. Infatti i giudici raffrontano gli accadimenti *sub iudice* verificatisi negli anni '60, con fatti simili accaduti negli anni '30 e sottolineano la differenza tra le zone prossime alla fabbriche e quelle distanti, così come evidenziano la differenza fra la situazione precedute e quella successiva all'impianto del depuratore della fabbrica, cercando di trovare una regolarità causale tra l'azione e l'evento in questione.

<sup>40</sup> Sono le parole del perito Augusto Murri, in *Quattro lezioni e una perizia*, Bologna, 1972, p. 127. Il perito fu chiamato a stabilire la relazione causale tra le percosse cagionate da un datore di lavoro al dipendente e la morte di questi per pneumonite, fatto accaduto nel 1884, conclusosi con una assoluzione per esclusione della relazione causale poiché la polmonite sorge frequentemente senza che sia preceduta da un trauma e perché il trauma spesso non porta alla polmonite. La soluzione fu criticata dal Murri, il quale evidenzia che non poteva essere raggiunta la prova del nesso causale in termini di *certezza*, ma essa doveva considerarsi raggiunta in termini di *probabilità*. Sulle parole di Murri, si veda il commento di riportate da F. Stella, *Leggi scientifiche, op. cit.*, p. 58.

<sup>41</sup> Nobili, *Nuove polemiche sulle cosiddette "massime di esperienza"*, *Riv. it. dir. proc. pen.*, 1969, p. 123 ss.

<sup>42</sup> M. Taruffo, *La prova del nesso causale, cit.*, p.101.

il suo impiego è compatibile con la funzione di garanzia della fattispecie<sup>43</sup>. In siffatta evenienza, è certo che il giudice possa spiegare il nesso di causalità ricorrendo a leggi scientifiche senza riproporre il linguaggio tecnico della scienza, anzi forse in alcuni casi ciò è auspicabile, posto che le sentenze devono essere comprensibili non solo al tecnico, ma anche all'uomo che di diritto non si occupa. Ne segue che il giudice può far ricorso alle massime di esperienza quando si limitano a semplificare l'enunciato di una legge scientifica causale. Per converso, sarebbe da escludere il richiamo di massime che siano espressioni di valutazioni etiche o culturali, ovvero di una massima di esperienza contraddetta da una legge scientifica, poiché ciò significherebbe affidare il giudizio alla "scienza privata" del giudice<sup>44</sup>.

Queste brevi battute sono necessarie per porre le basi delle nostre riflessioni in punto di causalità. Abbiamo investigato le ragioni per cui il metodo scientifico è l'unico su cui poter fare riferimento ai fini dell'accertamento del rapporto causale, per concludere nel senso che sono i principi di legalità e di responsabilità personale in materia penale a imporre al giurista di accogliere l'idea per cui il concetto di causa in ambito penale non può che essere legato "da un nesso inscindibile" con il concetto di causa accolto dal pensiero scientifico moderno<sup>45</sup>.

## 2. Le leggi scientifiche

Come detto, le leggi scientifiche sono in grado di *spiegare* la successione tra azione ed evento<sup>46</sup>. L'idea di fondo è che la scienza non si limita a descrivere dei fatti, come riteneva l'orientamento epistemologico affermatosi tra la seconda metà dell'Ottocento e i primi anni del Novecento<sup>47</sup>, ma tenta di spiegarli anche alla luce della descrizione dei

---

<sup>43</sup> La precisazione è di Stella, *op. cit.*, p. 150-152, il quale spiega tali affermazioni riportando l'esempio della spiegazione attraverso il senso comune del fenomeno di una sbarra di ferro che si allunga. In questo caso, la risposta "la sbarra è stata riscaldata" è frutto di un senso comune che condensa e assume tacitamente una legge scientifica, cioè la legge della dilatazione termica.

<sup>44</sup> M. Taruffo, *op. cit.*, pp. 104 e 105.

<sup>45</sup> F. Stella, *Leggi scientifiche, op. cit.*, p. 101.

<sup>46</sup> Sulla spiegazione causale nella scienza, E. Agazzi, *La spiegazione causale di eventi individuali (o singoli)*, in *Riv. it. dir. proc. pen.*, 1999, pp. 393 ss.

<sup>47</sup> Si tratta di un pensiero che prende le mosse dall'affermazione del pensiero positivista, secondo il quale lo scienziato deve osservare i fatti e analizzare i fenomeni e non proporre argomentazioni o ipotesi esplicative, come emblematicamente riassunto nel motto newtoniano "*hypotheses non fingo*". Nondimeno, a partire dalla fine del XIX sec. questa concezione epistemologica della scienza viene esasperata, fino al punto che prima si dissocia la spiegazione dei fatti dalla loro descrizione e poi si contrappone la prima alla seconda, per ritenere che la scienza genuina va alla ricerca di descrizioni e non spiega. Come osserva A.

dati, cercando di scoprire delle regolarità nel flusso degli eventi enucleando leggi generali utilizzabili a scopo di previsione<sup>48</sup>.

Siffatte leggi asseriscono una connessione tra un fatto e un evento, *spiegandola* sulla base di un rapporto di regolarità causale ovvero di probabilità relativa. Nel primo caso, le leggi scientifiche adoperate saranno quelle universali, nel secondo quelle statistiche. Entrambe queste leggi sono proprie del metodo scientifico di spiegazione dei fenomeni, per cui ciò che deve spiegarsi, cioè l'*explanandum* (E), è reso intellegibile poiché ricondotto a una classe di leggi (L) e di condizioni fattuali rilevanti (C), denominata *explanans*, secondo la terminologia impiegata da Hempel e Oppenheim<sup>49</sup>.

Le leggi universali postulano relazioni deduttive, che garantiscono che a date condizioni un certo fatto accade necessariamente<sup>50</sup> o ineccepibilmente, mentre le leggi statistiche si riferiscono a ipotesi probabilistiche e relazioni induttive<sup>51</sup>. Queste ultime sono espressione della versione più debole del modello scientifico<sup>52</sup> ma in molti campi bisogna farvi ricorso, poiché limitate sono le leggi universali di cui la scienza dispone. Prendendo atto di ciò, nel panorama epistemologico si è fatta avanti la tesi che riconosce valore esplicativo anche alle leggi statistiche e non solo a quelle universali<sup>53</sup>. Queste ultime compongono i processi esplicativi *nomologico-deduttivi*, terminologia che pone l'accento sulla circostanza che l'*explanans* contiene una o più leggi (L1, L2, ...) e una o più premesse o condizioni (C1, C2, ...) e sul fatto che il rapporto tra l'*explanans* e l'*explanandum* è di tipo deduttivo, poiché il secondo è connesso *necessariamente*, quindi

---

Pasquinelli, *op. cit.*, pp. 75 ss., si tratta di un punto di vista alquanto radicale, che riduce la scienza alla mera descrizione di fatti o di leggi e quindi non accettabile per vari motivi.

<sup>48</sup> A. Pasquinelli, *Nuovi principi di epistemologia*, Milano, Feltrinelli, 1974, p. 84.

<sup>49</sup> C. G. Hempel, P. Oppenheim, *Studies in the logic of explanation*, in *Philosophy of Science*, vol. 15, 1948, pp. 135-175, si veda in particolare p. 136, nota 2; C.G. Hempel, *Aspetti della spiegazione scientifica*, Milano, Il Saggiatore, 1987.

<sup>50</sup> W. H. Dray, *Filosofia e conoscenza storica*, Bologna, Il Mulino, 1973, p. 19.

<sup>51</sup> Nel ragionamento logico, gli argomenti vengono distinti in deduttivi e induttivi, riconoscendo i primi quando la relazione tra premesse e conclusione è tale che è assolutamente impossibile che le premesse siano vere a meno che la conclusione non sia anch'essa vera; i secondi, invece, sostengono soltanto che le premesse forniscono un qualche sostegno per quella conclusione; e allora, mentre gli argomenti deduttivi sono validi o invalidi, quelli induttivi possono essere valutati come migliori o peggiori, a seconda del grado di garanzia dato dalle premesse alle conclusioni. In argomento, I. M. Copi, C. Cohen, *Introduzione alla logica*, Bologna, Il Mulino, 1997, pp. 71 ss.

<sup>52</sup> W. H. Dray, *op. cit.*, p. 20.

<sup>53</sup> A favore del loro impiego si pronuncia C. G. Hempel, *Aspetti della spiegazione scientifica, cit.*, p. 106 ss., mentre a favore dell'utilizzo delle sole leggi universali si erano prima schierati Nagel, *La struttura della scienza, cit.*, p. 80, e Popper, *Logica della scoperta scientifica*, Torino, Einaudi, 1970, p. 14.

sempre senza eccezioni, a un insieme di condizioni empiriche antecedenti, sulla base delle leggi incluse nell'*explanans*<sup>54</sup>.

Le leggi statistiche rientrano nella spiegazione causale di tipo nomologico-induttivo o nomologico-statistico, dove la natura statistica della legge comporta che l'*explanans* (L e C) implichi l'*explanandum* soltanto con un adeguato grado di probabilità induttiva e non *necessariamente* o *ineccepibilmente*<sup>55</sup>. La forza dimostrativa del modello nomologico-deduttivo di derivazione hempeliana sarebbe quindi disomogenea, a seconda che si impieghino leggi universali o statistiche<sup>56</sup>.

Si è posta però la questione del carattere effettivamente “universale” delle leggi del primo tipo, posto che lo scienziato non può mai avere la certezza che una regolarità da lui individuata ricorra realmente in tutti i casi<sup>57</sup>.

Il fatto che le leggi universali non possono essere fondate sull’osservazione della totalità dei casi ha posto le basi per sostenere che esiste un’unica categoria di leggi scientifiche, che sono le “leggi probabilistiche”. L’affermazione mi pare condivisibile, ma è vero che la questione non è di cruciale rilevanza in materia penale nel momento in cui si ammette che il collegamento tra evento e azione possa essere stabilito con l’aiuto di leggi statistiche<sup>58</sup>. Piuttosto questo è un argomento per sostenere il carattere probabilistico dell’accertamento processuale.

Sul punto si osserva che la causalità scientifica richiamata nel diritto penale avrebbe natura probabilistica non solo per il carattere probabilistico tanto delle leggi universali quanto di quelle statistiche, ma anche per un’altra ragione, consistente nella necessità di ricorrere ad assunzioni tacite e alla clausola *ceteris paribus*, la quale è destinata a nascondere l’*impossibilità* di esplicitare le premesse del ragionamento<sup>59</sup>. Infatti, non è possibile per il giudice conoscere tutte le leggi e le condizioni necessarie per una spiegazione nomologico-deduttiva<sup>60</sup>, ma del resto, ciò non è possibile neanche per la scienza, la quale non possiede la conoscenza completa di tutte le leggi della natura. Si

---

<sup>54</sup> A. Pasquinelli, *op. cit.*, p. 96.

<sup>55</sup> *Idem*, p. 99 e 100.

<sup>56</sup> C. Piergallini, *Danno da prodotto e responsabilità penale*, Milano, Giuffrè, 2004, p. 161.

<sup>57</sup> Maiwald, *op. cit.*, p. 21; F. Mantovani, *Diritto penale, op. cit.*, p. 143; F. Stella, *op.cit.*, p. 308 e 309; M. Donini, *Il garantismo della condicio sine qua non e il prezzo del suo abbandono*, in *Riv. it. dir. proc. pen.*, 2011, p. 504 ss.

<sup>58</sup> Così F. Stella, *op. cit.*, pp. 311-312.

<sup>59</sup> O. Di Giovine, *Lo statuto epistemologico della causalità penale tra cause sufficienti e condizioni necessarie*, *Riv. it. dir. proc. pen.*, 2002, p. 652.

<sup>60</sup> F. Stella, *Leggi scientifiche, op. cit.*, p. 281.

pensi all'esempio del corvo proposto da Nagel. La legge di natura "tutti i corvi sono neri" comporta che se un dato uccello *u* è un corvo, allora è nero, tuttavia non spiega perché i corvi sono neri; ciò non significa che non sia una legge valida: essa attiene a un piano diverso, cioè non spiega perché tutti i corvi sono neri, ma spiega perché *u* è nero, dimostrando che il colore delle piume di *u* non è una peculiarità dello stesso, ma un tratto che *u* ha in comune con ogni altro uccello che, come lui, sia un corvo<sup>61</sup>. Pertanto la legge è valida perché riesce a spiegare un segmento causale, nonostante presupponga una tacita accettazione di certe condizioni (nell'esempio, perché i corvi sono neri). In molti casi, del resto, le scienze naturali (o le scienze impiegate nell'indagine storica) menzionano solo alcune delle condizioni *indispensabili* o *necessarie* di un dato avvenimento<sup>62</sup>. Ebbene, allo stesso modo il ragionamento del giudice presenta sempre delle assunzioni nomologiche e contingenti tacite, che non possono essere rese esplicite nel corso del processo e quindi rientrano tra le condizioni che si suppongono per date, cioè per esistenti o conosciute, assieme alle condizioni iniziali, che spesso sono non conosciute o soltanto azzardate<sup>63</sup>. Quindi il rapporto causale si riterrà provato solo in presenza di certe condizioni verificate e a parità di altre condizioni che sono non conosciute o soltanto azzardate (clausola *ceteris paribus*)<sup>64</sup> che impediscono di raffigurare la spiegazione causale a carattere deduttivo<sup>65</sup>.

Ne consegue che il rapporto causale è accertato all'esito di un processo che non è mai deduttivo, ma induttivo-probabilistico<sup>66</sup>, sia perché tutte le leggi naturali (quindi anche quelle che il giurista definisce universali) sono delle asserzioni probabilistiche<sup>67</sup>, sia per l'impossibilità di conoscere tutti gli antecedenti causali che operano

---

<sup>61</sup> E. Nagel, *La struttura della scienza*, Milano, Feltrinelli, 1968, p. 54. Nell'ambito della fisica, M. Planck, *La conoscenza del mondo fisico*, Torino, 1949, p. 120, secondo il quale "la fisica conosce, oltre alle leggi dinamiche, di rigorosa validità in tutti i casi particolari, anche ulteriori leggi che vengono dette statistiche, che hanno solo un carattere di probabilità ed in casi singoli ammettono eccezioni", utilizzando la terminologia leggi dinamiche (universali) e statistiche.

<sup>62</sup> E. Nagel, *op. cit.*, p. 575.

<sup>63</sup> F. Stella, *Leggi scientifiche, op. cit.*, p. 285.

<sup>64</sup> E. Nagel, *op. cit.*, p. 577, il quale nota che la clausola *coeteris paribus* è impiegata anche nella fisica. Riporta in proposito lo studio della traiettoria di un proiettile, che è può essere spiegata ricorrendo alla teoria newtoniana, pertanto si menzioneranno la latitudine cui il colpo è esploso, la resistenza dell'aria, la velocità iniziale del proiettile e altro ancora, ma probabilmente nella spiegazione non sarà menzionata la posizione della Terra rispetto alla propria galassia e a altri sistemi galattici, poiché la teoria newtoniana assume che la massa del proiettile sia costante ed è indipendente dalla velocità del proiettile stesso e dalla sua distanza da altri corpi (p. 577 in nota).

<sup>65</sup> F. Stella, *op. cit.*, p. 304.

<sup>66</sup> F. Stella, *op. cit.*, *Appendice*, p. 365-368; O. Di Giovine, *Probabilità statistica e probabilità logica nell'accertamento del nesso di causalità*, *Cass. pen.*, 2008, p. 2157.

<sup>67</sup> F. Stella, *op. cit.*, p. 311.

sinergicamente alla realizzazione dell'evento, da cui discende l'accettazione di condizioni iniziali e leggi che restano tacite<sup>68</sup>. In conclusione, il modello condizionalistico orientato secondo leggi scientifiche non può condurre alla certezza scientifica dell'accertamento, la quale resta un "miraggio utopico che cede il passo a una più ragionevole credibilità logica"<sup>69</sup>. Ciò, tuttavia, non stride con l'esigenza di certezza richiesta per l'attribuzione dell'evento alla condotta dell'imputato, atteso che nel giudizio penale la certezza va riferita al "*risultato finale* dell'operazione di sussunzione" che non deve "lasciare dubbi rispetto all'efficacia causale della condotta contestata all'accusato"<sup>70</sup> e tale certezza può derivare anche dall'impiego di generalizzazioni statistiche, purché si accerti che proprio quella regolarità si è concretizzata nel caso di specie<sup>71</sup> (accertamento effettuato, come si dirà a breve, attraverso la prova per esclusione).

Si comprende, allora, che l'utilità della legge scientifica, universale o statistica, risiede nella capacità della stessa di esprimere tra l'antecedente e il conseguente un rapporto che va oltre la semplice contingenza o la semplice successione cronologica tra i fenomeni, per cui si è soliti affermare che esse esprimono un "nesso più forte di una semplice concomitanza in via di fatto"<sup>72</sup>. Pertanto, una legge è scientifica se, oltre alla componente condizionale "se...allora", presenta una componente dotata di un potere esplicativo che spieghi la relazione tra antecedente e susseguente. Questo potere esplicativo è la conseguenza della *controllabilità* della legge attraverso degli esperimenti, ma anche del suo *carattere generale*, nel senso che possono essere applicate a un numero infinito di casi.

Controllabilità e generalità sono le caratteristiche delle leggi scientifiche secondo Stella, il quale però ritiene che la regola che emerge dall'enunciato della legge debba altresì godere di un *alto grado di conferma*<sup>73</sup>. Questo terzo requisito non è da tutti condiviso e, anzi, la giurisprudenza italiana si attesta pacificamente in senso contrario.

---

<sup>68</sup> O. Di Giovine, *Lo statuto epistemologico*, cit., p. 653.

<sup>69</sup> O. Di Giovine, *La causalità omissiva in campo medico-chirurgico al vaglio delle Sezioni Unite*, in *Foro it.*, II, 2002, pp. 608 ss., par. 4.

<sup>70</sup> F. D'Alessandro, *Le frequenze medio-basse e il nesso causale tra omissione ed evento*, in *Cass. pen.* 2007, p. 4828.

<sup>71</sup> Pertanto, se è vero che una probabilità prossima a 1 rende la legge statistica più valida, è anche vero che una percentuale elevatissima non è per ciò solo sufficiente per l'affermazione del nesso di condizionamento, che potrà essere smentito dalla concreta emersione di un diverso decorso causale, così M. Romano, *Commentario sistematico del codice penale, art. 40*, Milano, Giuffrè, 2004, p. 373.

<sup>72</sup> E. Nagel, *La struttura della scienza*, p. 57

<sup>73</sup> F. Stella, *Leggi scientifiche*, op. cit., pp. 123-126. Successivamente l'A. preciserà che deve trattarsi di un coefficiente probabilistico "vicinissimo a centro", *Idem, Giustizia e modernità*, p. 371 ss.

E infatti, una volta accettato l'impiego del metodo scientifico con leggi statistiche, la giurisprudenza e la dottrina si sono interrogate sul grado di probabilità minimo che deve essere richiesto al fine di ritenere provato il nesso di condizionamento. L'approdo di tale dibattito è noto: la giurisprudenza accetta oggi in modo pacifico il criterio della elevata *probabilità logica* o *logico-concreta*<sup>74</sup>, intesa come *alta credibilità razionale dell'accertamento*. Secondo questo approccio, che richiama il modello logico dell'abduzione<sup>75</sup>, non è necessario che la legge di copertura abbia un coefficiente di probabilità statistica elevato e prossimo alla certezza, poiché anche un basso grado di probabilità statistico-astratta può spiegare un caso concreto, in cui l'evento presenta un elevato grado di probabilità concreta di verifica, ma a condizione che sia dimostrata l'*assenza di fattori alternativi*<sup>76</sup> attraverso la *prova per esclusione*. Così ragionando, una frequenza statistica bassa *ex ante* non impedisce una spiegazione certa *ex post*: l'evento è spiegato statisticamente dalla legge di copertura e causalmente dall'accertamento a posteriori, escludendo altri possibili fattori causali<sup>77</sup> e in ciò consiste la probabilità logica.

E allora, dal confronto costante tra probabilità in astratto e probabilità in concreto, tre sono gli scenari prospettabili<sup>78</sup>: i due tipi di probabilità coincidono quando l'evento in concreto rientra nel novero dei casi che determinano l'alta probabilità della legge in astratto; oppure, l'evento non riesce ad essere spiegato dalla legge che abbia un elevato grado di probabilità statistico-astratta, poiché dei fattori causali alternativi ne escludono la sussunzione sotto quella legge di copertura, cosicché esso rientrerà tra i casi che fanno eccezione alla regola enunciata dalla legge statistica; infine, il caso può anche essere

---

<sup>74</sup> Si tratta dell'impostazione accolta da Sez. un., 10 luglio 2002, n. 30328, Franzese, *Cassazione Penale*, 2002, p. 3643, *Foro It.*, 2002 con nota di O. Di Giovine, p. 608, e da allora condivisa pacificamente dalla giurisprudenza successiva, tra cui, *ex multis*, Cass. pen., Sez. IV, 13 ottobre 2010, n. 43786, Cozzini, in particolare par. 5-7; Cassazione penale sez. IV, 22/02/2018, n. 22022, in *Cassazione Penale* 2019, 10, 3672.

<sup>75</sup> O. Di Giovine, *Probabilità statistica e probabilità logica*, cit., p. 2178, modello elaborato dal filosofo statunitense Charles Sanders Peirce.

<sup>76</sup> Il processo di esclusione dei decorsi alternativi era già stato valorizzato dalla sentenza sul disastro di Stava del 1990 (Cass. sez. IV, 6 dic. 1990, Bonetti, in *Foro it.*, 1992, II, p. 45, ove si legge: "il nesso di condizionamento tra azione ed evento potrà essere riconosciuto presente soltanto con una quantità di precisazioni e purché sia ragionevolmente – non certamente- da escludere l'intervento di un diverso processo causale". In dottrina già M. Donini, *La causalità omissiva e l'imputazione per l'aumento del rischio*, in *Riv. it. dir. proc. pen.*, 1999, pp. 48, 73; O. Di Giovine, *Lo statuto epistemologico*, op. cit., p. 654 e anche L. Eusebi, *Appunti sul confine tra dolo e colpa nella teoria del reato*, in *Riv. it. dir. proc. pen.*, 2000, p. 1057. Questa opinione sarà comunque criticata da F. Stella, il quale, commentando la pronuncia delle Sez. Un., affermerà che "è una questione di pura fortuna se il giudice riesce ad "azzeccare" l'individuazione di ciò che è realmente accaduto utilizzando frequenze medio-basse", *Verità, scienza e giustizia: le frequenze medio-basse nella successione di eventi*, *Riv. it. dir. proc. pen.*, 2002, p. 1236.

<sup>77</sup> M. Donini, *Il garantismo della condicio*, cit., pp. 511-512.

<sup>78</sup> F. Mantovani, *Diritto penale*, op. cit., p. 145.

inverso, allorché la probabilità in astratto che da una certa condotta derivi un determinato evento è bassa e cionondimeno al caso concreto è applicabile quella legge di copertura in quanto è provata l'assenza di fattori causali alternativi.

Quello proposto dalla giurisprudenza è un accertamento bifasico, che fa seguire all'individuazione di una legge di copertura e alla formulazione di un'ipotesi la prova che il decorso ipotizzato si è realizzato nel caso concreto, fermo restando che non è consentito “dedurre automaticamente dal coefficiente di probabilità statistica espresso dalla legge la conferma dell'ipotesi sull'esistenza del rapporto di causalità”<sup>79</sup>. Questo ragionamento, che persegue l'obiettivo di porre fine all'impiego giurisprudenziale di forme “deboli” di imputazione dell'evento, valorizza il processo come sede in cui si accerta la causa intesa nella sua nozione giuridica e non epistemologica<sup>80</sup>. Rinunciare al secondo momento dell'accertamento comporterebbe un'applicazione *de plano* della legge scientifica, la quale, invece, deve essere utilizzata come strumento. Siamo di fronte a un sapere scientifico che non rivela di per sé il nesso di causalità del caso oggetto del giudizio, ma che rappresenta la lente che il giurista deve indossare per leggere il legame eziologico del caso concreto. La legge di natura, allora, consentendo la previsione di eventi del tipo di quello verificatosi, è di per sé “muta” rispetto al singolo caso, poiché serve alla sola rappresentazione di un risultato ritenuto probabile, ma non consente di spiegare il singolo nesso; in altri termini, la legge scientifica, in astratto considerata, è espressione della causalità generale – in quanto consente la previsione di eventi del tipo di quello verificatosi<sup>81</sup>, che è un accertamento di cui il diritto penale non si accontenta, atteso che anche per il nesso di causalità non possono non valere “gli identici criteri di accertamento e di rigore dimostrativo che il giudizio penale riserva a tutti gli elementi costitutivi del fatto di reato”<sup>82</sup>.

---

<sup>79</sup> Cass, Sez. un. Franzese, *cit.*, par. 8.

<sup>80</sup> In questo senso, M. Donini, *op. cit.*, p. 81.

<sup>81</sup> In questo senso, O. Di Giovine, *Lo statuto epistemologico*, *op. cit.*, p. 659.

<sup>82</sup> S.U. Franzese, *cit.*

### 3. La causalità generale e l'aumento del rischio

Il concetto di causalità generale è contrapposto a quello di causalità individuale e tali espressioni linguistiche rispecchiano l'accertamento bifasico proposto a partire dalla sentenza Franzese<sup>83</sup>.

Si è visto che la causalità generale si limita a fornire un'informazione circa l'attitudine di un evento a determinare l'aumento o la diminuzione delle probabilità del verificarsi di un altro evento<sup>84</sup> considerato *ex ante*, investigando il prodursi del fenomeno *in astratto*, mentre nulla riferisce sull'evento *concreto*. Essa postula l'impiego di leggi scientifiche che rappresentano un risultato futuro ritenuto probabile o seriamente possibile, ma non spiega il singolo nesso, poiché si limita alla formulazione di una ipotesi. Affinché il nesso possa dirsi dimostrato è necessario verificare e confermare l'attendibilità dell'ipotesi nel caso concreto. Ne deriva che la causalità generale, isolatamente considerata, è allora una causalità insufficiente per il diritto, poiché basare la verifica del nesso di condizionamento solo sulla causalità generale significa dare rilevanza alla mera idoneità della condotta di aumentare il rischio che l'evento si verifichi.

Il problema dell'ingresso della teoria dell'aumento del rischio per l'accertamento del nesso di condizionamento è già stato oggetto di analisi con riferimento a fenomeni di disastri, di malattie, di danni da prodotto, cioè di situazioni empirico-criminose collocate su grandezze di scala, difficili da governare con le tradizionali categorie penalistiche, ma più facili da dominare col diritto penale del rischio<sup>85</sup>.

Questi temi si ripropongono con evidenza anche nel settore dell'intelligenza artificiale, tecnologia che sottopone il diritto (non solo penale) a un ulteriore "shock da modernità"<sup>86</sup>, che si aggiunge a quello che ormai da tempo è causato dalla "civilizzazione tecnica" e dal progresso tecnico-scientifico.

---

<sup>83</sup> Questa terminologia è utilizzata per descrivere il ragionamento proposto dalle Sez. Un. Franzese, ta gli altri, da R. Bartoli *op. cit.*, pp. 31 ss.; G. De Vero, *Il nesso causale e il diritto penale del rischio*, in *Riv. it. dir. proc. pen.*, 2, 2016, p. 677.

<sup>84</sup> F. Stella, *Il contesto*, *cit.* p. 10.

<sup>85</sup> Sul punto, si rimanda alla chiara esposizione di G. De Vero, *Il nesso causale e il diritto penale del rischio*, in *Riv. it. dir. proc. pen.*, 2, 2016, p. 670 ss.

<sup>86</sup> F. Stella, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, 3 ed. Milano, Giuffrè, 2003, p. 293.

Un altro aspetto merita attenzione: la causalità generale postula l'impiego di *leggi scientifiche*, cioè leggi *causali*, che spiegano l'insorgenza dell'*evento tipo* e non di quel singolo evento, da qui appunto la sua generalità<sup>87</sup>.

Non è però cosa semplice stabilire il ruolo che alcune conoscenze scientifiche possano avere in sede di verifica della causalità generale. Ad esempio Stella esclude l'impiego di scienze nuove, quali l'epidemiologia, la tossicologia, la medicina del lavoro, la medicina della valutazione del rischio<sup>88</sup>, poiché riferite non all'individuo, ma a gruppi di popolazioni. Si tratta di scienze che delineano l'idoneità di un dato fattore ad aumentare il rischio della verificazione di un certo evento spesso connesso all'esercizio di attività espletate in un contesto organizzativo complesso, nel quale non è agevole dimostrare l'esistenza del nesso di condizionamento in modo rigoroso<sup>89</sup>. Questi studi offrono conoscenze fondate "non su leggi causali, ma su dati statistici" e si riferiscono a gruppi di popolazioni piuttosto che a individui; il loro impiego sarebbe allora da escludere poiché "l'assenza di una legge causale impedirebbe la sussunzione degli accadimenti concreti sotto la norma giuridica astratta che individua la causalità come requisito di fattispecie"<sup>90</sup>.

Altri autori, invece, ritagliano anche a queste scienze un ruolo all'interno della causalità generale<sup>91</sup>.

Lo studio di questa categoria merita allora approfondimento, poiché nel contesto di nostro interesse le conoscenze scientifiche ci offrono dei dati statistici, cioè i dati del test di validazione dei sistemi di IA, di cui è bene investigare il possibile ruolo nell'accertamento del nesso di causalità.

### **3.1 La probabilità nelle tecnologie di IA: margine di errore e *accuracy* come strumenti a servizio della decisione giudiziaria**

Come si è visto dallo studio condotto nel primo capitolo, il test di validazione dei sistemi di IA viene effettuato per verificare le prestazioni della tecnologia che si vuole

---

<sup>87</sup> In questo senso, O. Di Giovine: "il richiamo alle leggi causali (...) si riduce a strumento che consente la previsione di eventi *del tipo* di quello verificatosi (causalità -inevitabilmente- generale)", in *Lo statuto epistemologico*, *op. cit.*, p. 659.

<sup>88</sup> F. Stella, *Giustizia e modernità*, *cit.*, p. 293.

<sup>89</sup> A. Gargani, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato di fronte alle esigenze di controllo penale delle nuove fenomenologie di rischio*, in *LP*, 2, 2011, p. 400.

<sup>90</sup> F. Stella, *Giustizia e modernità*, *cit.*, p. 237.

<sup>91</sup> Si pensi ad esempio, al ruolo preponderante che l'epidemiologia ha nella tesi dell'accertamento alternativo, *infra*, par. 3.2.

immettere nel mercato. Da tale test si ricavano dei dati statistici rappresentativi del livello di accuratezza della decisione algoritmica<sup>92</sup> e del margine di errore, il quale, va ricordato, è un aspetto immanente a certe tecnologie di IA, atteso che un errore pari a zero è una condizione ideale, ma impossibile in contesti complessi. Si tratta di dati che possono poi essere approfonditi nei loro risultati attraverso la matrice di confusione, che permetterà di individuare gli errori di classificazione e le risposte corrette all'interno di ciascuna classe di elementi.

A questo punto potremmo chiederci se questi valori di cui dispone il produttore, estrapolati in sede di validazione, possano avere un ruolo nel momento dell'accertamento del nesso di causalità e, se sì, quale.

Per discutere di questo aspetto bisogna prima dare per presupposto che tali dati siano accessibili. Su questo punto potrebbe rappresentare un ostacolo il segreto industriale, che tutelerebbe il produttore. Si tratta di una questione che andrebbe affrontata sul piano legislativo e a livello europeo. Il percorso sembra essere tracciato dal legislatore europeo in favore della trasparenza e dell'esplicabilità di questi sistemi. Tali principi, che secondo la visione europea devono guidare la produzione dei sistemi di IA, fungono da argine a forme di *opacità giuridica* alimentate dal segreto industriale, e dovrebbero tradursi in regole applicative puntuali e precise, in grado di bilanciare l'interesse alla produzione con quello alla conoscenza del funzionamento di questi sistemi.

La conoscibilità di tali informazioni le renderà certamente fruibili in sede processuale, ma con quale funzione? Riflettiamo.

Si tratta di conoscenze che provengono dalle discipline informatiche, quindi una scienza diversa da quella giuridica. Queste informazioni utilizzano un linguaggio statistico, fornendo delle percentuali circa il margine di errore e l'accuratezza della decisione del sistema di IA. Tali dati sono estrapolati in sede di test del sistema, quindi durante il processo produttivo, in cui si simula una ideale condizione di utilizzo dello stesso. In sostanza si tratta di esperimenti, che quindi sono ripetibili nel tempo e verificabili, i cui esiti hanno un carattere generale, poiché applicabili per ogni utilizzo del sistema di IA. Tutto ciò è sufficiente per affermare che tali dati sono delle leggi scientifiche, così come prima definite, che stabiliscono il legame in termini probabilistici tra una data azione e un dato evento? Probabilmente no, perché i dati del test di

---

<sup>92</sup> Accuratezza che, ricordiamo, è il parametro riferito ai compiti di classificazione, quelli ai quali in questo contesto abbiamo circoscritto lo studio per ragioni di sintesi e semplificazione.

validazione non sono costruiti *ab origine* in modo causale, non riferiscono una correlazione causale, ma mere *correlazioni statistiche*.

Sono correlazioni statistiche i dati epidemiologici, di cui come si è poco sopra ricordato, autorevole dottrina ha escluso l'attribuibilità alla categoria delle leggi statistiche, trattandosi piuttosto di elementi di fatto utilizzabili come base o come elemento di controllo di ipotesi scientifiche<sup>93</sup>. Un parallelo può allora essere tracciato con i dati del test di validazione. Tali dati forniscono informazioni su situazioni di fatto già oggetto di esperimento, che possono essere trasposte su situazioni di fatto future e quindi potenziali: con essi è possibile sapere, per ritornare a un esempio già proposto, che sottoponendo al sistema di classificazione una serie di quadri rappresentanti figure femminili, questo potrà riconoscere il dipinto di Vermeer, *La ragazza col turbante*, con una certa frequenza statistica.

Se consideriamo questo compito di classificazione calato in un sistema di guida autonoma, allora significa che si dispone dell'informazione relativa alla frequenza con la quale il pilota automatico riconosce un ostacolo o individua una certa segnaletica orizzontale o verticale. Se il pilota automatico non si arresta in presenza di un semaforo rosso o di un passaggio pedonale, i dati del test di validazione ci sapranno dire con quale frequenza il sistema errava nel riconoscimento di segnaletiche dello stesso tipo ovvero con quale frequenza le riconosceva con successo.

Si capisce, quindi, come i dati del test di validazione non spieghino la correlazione diretta con l'evento penalmente rilevante, semmai essi descrivono la relazione tra operatività del sistema e decisione cui il sistema addiviene. Potrà quindi sostenersi che nel momento in cui tale decisione incide sulla produzione dell'evento, i dati del test di validazione potranno avere un certo valore nell'accertamento del nesso di condizionamento. In particolare, tali informazioni emergenti dal test di validazione unitamente ad altri elementi (e non isolatamente considerati) forniscono al giudice un'ipotesi scientifica sulla potenzialità lesiva del sistema di IA, utilizzabile per comprendere quale sia il contributo dato dal prodotto intelligente (e quindi del produttore) e dal conducente (utilizzatore del sistema) nella realizzazione di un dato evento, quale può essere le lesioni riportate dal pedone.

---

<sup>93</sup> D. Pulitanò, *op. cit.*, p. 792; A. Kaufmann, *Tatbestandmässigkeit und Verursachung im Contenganverfahren*, in *JZ*, 1971, p. 569, richiamato da F. Stella, *Giustizia e modernità, cit.*, p. 237.

Ciò significa che l'impiego di tali dati in sede giudiziale va circoscritto, poiché da soli non bastano per affermare la portata condizionante della scelta del sistema di IA nella causazione del danno, e dovranno appunto servire a individuare un'ipotesi che poi dovrà essere confermata da una teoria scientifica<sup>94</sup>. Giungendo alla conclusione opposta si arriverebbe all'assurdo di ritenere che l'esistenza di un dato indicante il margine di errore che il sistema è portato a commettere (*accuracy*) sia di per sé sufficiente a provare la causalità, con la conseguenza di dare spazio all'aumento del rischio in sede di accertamento del nesso eziologico.

Ben si comprende, allora, come il settore dei danni da uso di sistemi di IA si presti a forme di accertamento della causalità che si accontentano del solo aumento del rischio della verificazione dell'evento, così cadendo nella tentazione di ricorrere a teorie causali ormai superate dopo la sentenza Franzese. Infatti, valorizzare l'aumento del rischio significa rinunciare ad approfondire il secondo momento dell'accertamento causale bifasico, volto a ricercare la causalità individuale o specifica, accontentandosi della mera idoneità della condotta ad aumentare il rischio di verificazione dell'evento, in positivo, ovvero, in negativo, della diminuzione delle *chances* di mancata verificazione dell'evento, così superando la dimensione naturalistica della causalità a favore di un modello interamente normativo, che piuttosto che accertare l'evento, tende a stigmatizzare la violazione di una regola prudenziale<sup>95</sup>.

Le questioni che vengono in rilievo non sono nuove e si ricollegano alla "crisi del diritto penale d'evento" in favore dell'emersione del "diritto penale del rischio" in rapporto a fenomenologie di danno correlate ad attività a base lecita<sup>96</sup>, quale è appunto l'attività produttiva e, più in generale, l'attività di impresa. Proprio la responsabilità dell'imprenditore ha rappresentato e tutt'oggi rappresenta un orizzonte problematico per l'accertamento della causalità, atteso che la tutela della vita e dell'integrità fisica, che entra in gioco in un contesto di attività (spesso pericolosa) consentita, spinge verso l'affermazione di forme di responsabilizzazione penale anticipata. In particolare si è molto discusso intorno alla responsabilità del datore di lavoro in ordine al mancato adempimento di obblighi di sicurezza e di prevenzione di malattie professionali, che ha sollecitato un ampio dibattito dottrinale sull'incertezza degli effetti patogeni

---

<sup>94</sup> In questo senso, ma con riferimento ai risultati di indagini epidemiologiche, D. Pulitanò, *op. cit.*, p. 792.

<sup>95</sup> C. Piergallini, *op. cit.*, p. 177.

<sup>96</sup> A. Gargani, *op. cit.*, p. 398.

dell'esposizione a sostanze quali l'amianto<sup>97</sup>. Questo contesto ha favorito l'approfondimento del contributo che all'accertamento causale può dare l'epidemiologia e le riflessioni maturate in relazione a tale fenomenologia di danni mi sembrano molto pertinenti nel contesto del nostro studio perché l'epidemiologia, da un lato, e i risultati del test di validazione dei sistemi di IA, dall'altro, mi sembrano in parte sovrapponibili, come tenterò di spiegare di seguito. La nostra indagine si rivolgerà allora alle conclusioni che una parte della dottrina ha tratto con riguardo all'impiego delle conoscenze epidemiologiche e si cercherà di comprendere se esse siano estensibili anche ai risultati del test di validazione. Ma un avvertimento è sin da ora d'obbligo: mentre intorno all'utilizzabilità delle indagini epidemiologiche si è discusso con riguardo a casi in cui era in concreto incerta la causa di insorgenza di una patologia, quindi dell'offesa al bene giuridico tutelato, per i casi di danni da sistemi di IA non vi sarà sempre tale incertezza. Anzi, l'esclusione di problemi di funzionamento dell'hardware e di altre cause dell'evento ci porterà a ritenere che il difetto del prodotto risiede nel software, nell'algoritmo al quale i dati di validazione si riferiscono. Si tratta di una conclusione cui si giunge escludendo fattori causali alternativi che, pertanto, non può essere basata sul solo dato che attesti un margine di errore potenziale del sistema, cioè su quella informazione fornitaci dal test di validazione.

È venuto quindi il momento di occuparci di epidemiologia, senza pretese di completezza, poiché altro è l'oggetto del nostro studio e essa rappresenta solo un mezzo per leggere un fenomeno nuovo, non ancora oggetto di trattazione nelle aule dei tribunali o delle corti italiane.

### **3.2 Indagine epidemiologica e risultati del test di validazione dei sistemi di IA: (tante) analogie e (poche) differenze**

Nell'alveo della causalità generale si colloca l'epidemiologia, scienza medica che studia il ritmo con cui si manifestano le malattie e le condizioni che favoriscono o ostacolano il loro sviluppo<sup>98</sup>. Una prima componente dell'epidemiologia è quella

---

<sup>97</sup> Tra i numerosi contributi, anche F. Palazzo, *Morti da amianto e colpa penale*, *Dir. pen. proc.*, 2011, pp. 186 ss.; R. Bartoli, *La responsabilità penale da esposizione dei lavoratori ad amianto. Un'indagine giurisprudenziale*, [www.penalecontemporaneo.it](http://www.penalecontemporaneo.it); L. Maserà, *Danni da amianto e diritto penale*, [www.penalecontemporaneo.it](http://www.penalecontemporaneo.it); S. Zirulia, *Esposizione a sostanze tossiche e responsabilità penale*, Milano, Giuffrè, 2018, il quale offre un'analisi dettagliata delle pronunce sui *toxic cases* in Italia; D. Pulitanò, *Gestione del rischio*, *op. cit.*

<sup>98</sup> *Dizionario di medicina*, Treccani, 2010, voce epidemiologia, [www.treccani.it](http://www.treccani.it).

*descrittiva*, che indica la frequenza di malattie o condizioni o esposizioni di interesse sanitario; una seconda componente è l'epidemiologia *eziologica*, che indaga la relazione esistente tra un'esposizione (cioè tutto ciò che può avere un effetto sulla salute, come un trattamento o un fattore di rischio) e la frequenza di malattie<sup>99</sup>. Mentre l'epidemiologia descrittiva spiega che l'insorgenza di malattie non è casuale ed è utile per stabilire strategie di prevenzione, quella eziologica individua i possibili fattori di rischio nell'insorgenza di malattie.

Per quantificare la frequenza del verificarsi di un evento, l'epidemiologia si avvale di tre misure: il *rischio*, che riferisce la probabilità che l'evento si verifichi; la *prevalenza*, che corrisponde alla porzione della popolazione affetta dalla malattia; l'*incidenza*, che si riferisce alla frequenza con la quale nuovi eventi di malattia si verificano in una popolazione a rischio<sup>100</sup>.

Il risultato di uno studio epidemiologico fornisce delle stime di misure di frequenza di eventi, cioè l'*incidenza* e la *prevalenza*, e stime di misure di frequenza di effetto, cioè il rischio relativo e il rischio attribuibile.

La dottrina penalistica inquadra l'epidemiologia nella causalità generale perché produce dati relativi a *gruppi di popolazioni*, di cui poi si tiene conto nell'esame medico di singoli soggetti<sup>101</sup>. Accade, cioè, che attraverso lo studio di popolazioni si è in grado di prevedere, con un margine di incertezza predefinito, cosa può avvenire a livello del singolo soggetto<sup>102</sup>.

Poiché l'epidemiologia fa uso della statistica, i problemi che si pongono sono gli stessi che si osservano nelle tecnologie di IA che applicano un metodo statistico basato sui dati.

In particolare, anche le indagini epidemiologiche possono presentare dei *bias*<sup>103</sup>, cioè errori sistemici che possono portare a stime errate, ed è sempre necessario indicarne l'*accuratezza* o *validità* oltretutto il *grado di incertezza* o *precisione*<sup>104</sup>.

---

<sup>99</sup> G. Traversa, *Evidenza epidemiologica e decisioni nelle aule di tribunale*, in *Scienza, diritto e processo penale nell'era del rischio*, A. Amato, G. Flora, C. Valbonesi (a cura di), Torino, Giappichelli, 2019, p. 259.

<sup>100</sup> R. S. Greenberg, *op. cit.*, p. 47 e 48.

<sup>101</sup> F. Stella, *Giustizia e modernità*, *op. cit.*, pp. 294 ss.

<sup>102</sup> G. Traversa, *op. cit.*, p. 260; R. S. Greenberg, *Medical Epidemiology. Population Health and Effective Health Care*, 5 ed., McGraw-Hill Education, 2015, p. 27.

<sup>103</sup> R. S. Greenberg, *op. cit.*, p. 433.

<sup>104</sup> G. Traversa, *op. cit.*, p. 270.

Queste somiglianze tra l'epidemiologia e il procedimento decisionale di un sistema di IA basato sul *machine learning* sono a questo punto utili per determinare la rilevanza delle conoscenze che si hanno sul sistema intelligente nello stabilire il nesso causale.

In altri termini, la questione attiene al ruolo che possono avere i parametri di accuratezza, precisione e sensibilità, noti al produttore e possibilmente anche all'utilizzatore del sistema di IA, nel processo di accertamento del legame eziologico.

Prendiamo l'esempio di un software impiegato nel sistema della giustizia per coadiuvare la decisione del giudice sul rischio di recidiva di un certo soggetto. Sappiamo che il software al momento dell'acquisto ha un'accuratezza elevata pari al 90%, con un margine di errore che si attesta intorno al 10%. Questi valori ci possono essere utili per stabilire se nel caso concreto la risposta fornita dal software è corretta o meno? Ovvero, possiamo affermare che la risposta fornita rientra nella percentuale di casi in cui il software commette un errore?

Consideriamo a questo punto il caso di un'indagine epidemiologica che dimostri che l'esposizione continua a una data sostanza, ad esempio l'amianto, comporti nel 10% dei casi l'insorgere di una data malattia, nell'esempio, mesotelioma<sup>105</sup>. Possiamo affermare che l'insorgere della malattia nel soggetto X, che si esponeva per un lungo periodo di tempo alla sostanza di cui allo studio epidemiologico, sia dovuto all'esposizione alla sostanza tossica? Ovvero, possiamo dire che la vicenda del soggetto X rientri in quel 10% di casi in cui dall'esposizione alla sostanza deriva la malattia?

Come si è visto prima, la dottrina maggioritaria e la giurisprudenza recente concordano nell'escludere l'impiego esclusivo dei risultati epidemiologici per stabilire la portata condizionante di una data condotta, motivando alla luce del carattere generale di siffatta causalità che indaga classi di popolazioni, cioè eventi astratti, e non contingenze particolari o fatti concreti. I risultati degli studi epidemiologici possono solo rappresentare un riferimento generale nell'ambito della causalità, perché individuano *fattori di rischio* e non cause<sup>106</sup> e le patologie di riferimento sono analizzate come fenomeni di serie o di

---

<sup>105</sup> Come osserva S. Zirulia, *op. cit.*, pp. 71 ss, la legge scientifica sulla riconducibilità causale del mesotelioma all'amianto ha un coefficiente statistico medio-basso, del 5-10%; ciononostante, i possibili fattori di rischio alternativi sono pochi e molto rari e per questo l'esclusione di questi ultimi è un momento semplice della decisione giudiziaria. La difficoltà è piuttosto dovuta alla lungolatenza della malattia, che comporta che quasi sempre la persona offesa sia stata sottoposta nel corso della sua vita a diverse esposizioni, quindi dei problemi si pongono in ordine all'efficacia causale delle esposizioni successive alla prima.

<sup>106</sup> M. Lotti, *Mesotelioma: scienza e congetture nel processo penale*, in *Processo alla scienza*, R. Borsari (a cura di), Padova, Padova University Press, 2016, p. 24.

massa, non di eziologie individuali<sup>107</sup>; ne consegue che l'epidemiologia "serve, ma non basta", in quanto non risulta provata la causalità individuale, cioè la derivazione causale della singola patologia, poiché per definizione l'epidemiologia, scienza di popolazioni, non è in grado di spiegare tale nesso<sup>108</sup>: essa si occupa di previsione di tipi di evento, non di spiegazione di singoli eventi<sup>109</sup>.

Altri hanno precisato che il vero limite dell'epidemiologia non è quello di essere riferita a popolazioni, ma il fatto di prendere in considerazione fenomeni la cui eziologia è in larga misura ignota poiché la conoscenza di tutte le cause che portano a una patologia manca e dunque non è possibile escludere decorsi causali alternativi<sup>110</sup>. Così la dottrina ritiene che le rilevazioni epidemiologiche non sono leggi scientifiche nella forma di leggi statistiche, trattandosi piuttosto di *correlazioni statistiche* che forniscono informazioni sui fattori di rischio che comportano l'insorgenza di una data malattia<sup>111</sup>. Per questo motivo non possono da sole essere alla base di un accertamento causale, ma possono al più servire a corroborare un nesso già messo in evidenza da un'altra legge probabilistica che offra una spiegazione individuale che l'epidemiologia potrà al più supportare, ma non determinare<sup>112</sup>.

Tuttavia, bisogna dare atto della presenza di un'opinione diversa, allo stato minoritaria, che valorizza i dati epidemiologici per la definizione della correlazione causale. Il riferimento è fatto alla teoria dell'accertamento alternativo dell'evento per quanto concerne l'individuazione della vittima (*Opfer-Wahlfeststellung*) di matrice tedesca, con la quale si addivene a una sentenza di condanna del soggetto che ha

---

<sup>107</sup> M. Donini, *Il garantismo della condicio*, cit., p. 513.

<sup>108</sup> L. Masera, *Ancora sulla qualificazione penalistica dell'evidenza epidemiologica perché anche nella società del rischio è legittimo il ricorso al diritto penale d'evento*, in G. De Francesco, G. Morgante (a cura di), *Il diritto penale di fronte alle sfide della "società del rischio"*, Torino, Giappichelli, 2017, p. 390; C. E. Paliero, *Causalità e probabilità tra diritto penale e medicina legale*, *Riv. It. med. Leg.*, 4, 2015, p. 1515

<sup>109</sup> O. Di Giovine, *Ancora sull'infezione da sars-cov-2: omicidio e lesioni personali tra incertezze scientifiche, sovradeterminazioni causali e trappole cognitive*, *Leg. Pen.*, 27.01.2021, p. 5.

<sup>110</sup> R. Bartoli, *Il problema della causalità penale*, Torino, Giappichelli, 2010, p. 69.

<sup>111</sup> D. Pulitanò, *Gestione del rischio*, cit., p. 792. Sono invece considerate leggi statistiche da M. Donini, il quale però ne evidenzia la flebile rilevanza causale, giungendo sostanzialmente alle stesse conclusioni di D. Pulitanò, -quantomeno con riferimento ai reati di evento di danno- ritenendo che "nessuna spiegazione individuale è possibile sulla base di tali leggi, se non a supporto di altri criteri di valenza causale", ma, precisa l'A, "l'impiego di tali leggi, però, è possibile nell'imputazione di eventi di pericolo, là dove siano in gioco popolazioni (e dunque 'classi') di individui, come nei reati contro la salute pubblica, che sono reati "di comune pericolo" [...] dove cioè non è necessario individuare singoli soggetti passivi e stabilire l'incidenza del pericolo sulla loro salute individuale". In sostanza, la proposta dell'A. è quella di applicare a un contesto di pericolo comune (e non di danno) la tesi dell'accertamento alternativo proposta (per la quale v. *infra*), così da circoscrivere la rilevanza causale delle leggi epidemiologiche all'accertamento degli eventi di comune pericolo. M. Donini, *Il garantismo della condicio*, cit., p. 513-515.

<sup>112</sup> *Ibidem*, p. 792; C. Piergallini, *Danno da prodotto*, cit., pp. 228-230.

sicuramente causato l'offesa, nonostante rimanga ignota l'identità della vittima<sup>113</sup>. In questo contesto, la prova epidemiologica dimostrerebbe che l'esposizione a sostanze tossiche è la *condicio sine qua non* di un certo numero di morti in eccesso, che senza l'esposizione non ci sarebbero state; tale numero coincide col valore del c.d. *rischio attribuibile*, dato dalla differenza tra le patologie registrate nella classe degli esposti alla sostanza tossica e quelle rilevate nella classe dei non esposti, che consentirebbe di accertare la causalità individuale in ordine al numero di offese cagionate dalla condotta, pari alla quota di eventi che lo studio epidemiologico ricollega all'effetto della sostanza.

Ferme le riserve cui tale opinione si espone<sup>114</sup>, relative principalmente a un ritorno alla teoria dell'aumento del rischio e all'impossibilità di individuare nominativamente le vittime, come invece è richiesto dal diritto penale d'evento inteso non in astratto, ma in concreto, il problema nell'ambito delle tecnologie di IA mi pare sia rappresentato anzitutto dalla difficoltà di individuare le due classi di soggetti che poi saranno messe a confronto, secondo il metodo tipico degli studi epidemiologici di coorte, in cui si confronta la classe degli "esposti" a una data sostanza (ad esempio i lavoratori di una data impresa) con quella dei "non esposti". Nel contesto di nostro interesse, si dovrebbe probabilmente distinguere tra "utilizzatori" e "non utilizzatori" della tecnologia di IA coinvolta nella produzione dell'evento, per poi isolare all'interno della prima categoria la classe delle vittime innominate, ottenuta attraverso il calcolo del valore del rischio attribuibile. Nondimeno, la categoria degli utilizzatori del prodotto è una classe troppo generica e di difficile individuazione se si considera che la tecnologia intelligente può essere inserita in vari beni di cui costituisce un elemento che verrà poi a comporre il prodotto finito. Ad esempio, la tecnologia di riconoscimento di immagini può essere

---

<sup>113</sup> L. Masera, *Accertamento alternativo e evidenza epidemiologica nel diritto penale. Gestione del dubbio e profili causali*, Milano, Giuffrè, 2007; L. Masera, *Evidenza epidemiologica di un aumento di mortalità e responsabilità penale*, *Riv. trim. dir. pen. cont.*, 3-4, 2014, pp. 343-365; La tesi è ritenuta "convincente" da O. Di Giovine, *La causalità tra scienza e giurisprudenza*, *Riv. it. med. Leg.*, 1/2016, p. 42.

<sup>114</sup> Sul dibattito suscitato in dottrina da tale tesi, si rinvia a F. Bartoli, *Il problema della causalità penale*, *op. cit.*, p. 82-85; M. Donini, *Il garantismo della condicio sine qua non*, *cit.*, p. 514 ss. e *ivi* la declinazione della tesi dell'accertamento alternativo proposta dall'autore con riferimento ai reati di comune pericolo (già accennata *supra* in nota); A. Manna, *I rapporti tra sapere scientifico e sapere giudiziario*, in *Cass. pen.*, 2009, p. 3636; F. Viganò, *Il rapporto di causalità nella giurisprudenza penale a dieci anni dalla sentenza Franzese*, *Riv. Trim. dir. pen. cont.*, 3, 2013, p. 380 ss; F. D'Alessandro, *Art. 40*, in *Codice penale commentato*, E. Dolcini, G. Marinucci (a cura di), III ed., 2011, p. 411 ss. Le osservazioni critiche sono state affrontate dall'autore in scritti successivi, v. in particolare, L. Masera, *Evidenza epidemiologica di un aumento di mortalità*, *op. cit.*, e *Id.*, *Ancora sulla qualificazione penalistica dell'evidenza epidemiologica perché anche nella società del rischio è legittimo il ricorso al diritto penale d'evento*, in *Il diritto penale di fronte alle sfide della "società del rischio"*, G. De Francesco, G. Morgante (a cura di), Torino, Giappichelli, pp. 389 ss.

inserita in un'auto a guida autonoma e potrebbe essere venduta a molteplici aziende automobilistiche e non a una sola; allo stesso tempo essa potrebbe essere inserita in sistemi di diagnostica di malattie o di riconoscimento biometrico, e potrebbe persino rappresentare anche una piccola frazione della tecnologia in uso da parte di un sistema di IA più complesso. Si comprende quanto sia difficoltoso in questi casi individuare in modo esatto il numero degli utilizzatori della tecnologia per poi circoscrivere al suo interno il numero di eventi lesivi causalmente associati all'utilizzo di quella tecnologia. Specularmente, è altrettanto difficile individuare la categoria di soggetti prima definita dei "non utilizzatori": dovremmo forse riferire tale categoria a tutti gli utilizzatori di prodotti simili ma che non fanno uso di tecnologie di IA, o considerare anche i prodotti che impiegano tecnologie differenti da quella in questione?

Queste difficoltà non consentono di preparare il terreno per lo sviluppo della teoria dell'accertamento alternativo, che presuppone che sia provata oltre ogni ragionevole dubbio la colpevolezza dell'imputato, nonostante permangano incertezze in ordine alla ricostruzione del fatto costituente reato. Orbene, nel caso dell'epidemiologia si sostiene che "se uno studio scientificamente affidabile afferma che l'esposizione ad un fattore di rischio ha cagionato un aumento di mortalità, ciò significa che un certo numero di persone sono morte in ragione di tale esposizione"<sup>115</sup>; nel settore dei danni da produzione e uso di tecnologie di IA la certezza in ordine alla responsabilità dell'imputato è difficile da raggiungere attraverso lo stesso ragionamento proposto per l'impiego degli studi epidemiologici.

In sintesi, mi pare che a questa tesi non possa allo stato farsi ricorso per gli accertamenti causali in materia di danni da produzione e utilizzo di sistemi di IA e le ragioni mi sembrano vicine a quelle che in dottrina sono state avanzate nel rilevare come la tesi dell'accertamento alternativo non sia spendibile per la verifica della rilevanza causale della condotta umana nella contrazione del virus Covid-19, dove forte è l'incertezza numerica della classe degli esposti e dei non esposti al rischio della malattia, sia perché le indagini sono compiute "a campione", sia perché la presenza di positivi era sottostimata a causa dell'a-sintomaticità della manifestazione del virus<sup>116</sup>.

---

<sup>115</sup> L. Masera, *Evidenza epidemiologica di un aumento di mortalità*, *op. cit.*, p. 357.

<sup>116</sup> O. Di Giovine, *Ancora sull'infezione da sars-cov-2*, *op. cit.*, par. 6.; M. Di Florio, *Causalità e colpa medica in rapporto all'emergenza sanitaria*, *Leg. pen.*, 17.09.2021, p. 4; G. M. Marzano, *Brevi note sulla prova della causalità nel contagio da Covid-19*, *Cass. pen.*, 2020, pp. 3114-15.

Fermo quanto brevemente osservato in ordine alla teoria dell'accertamento alternativo, possiamo ritenere che le conclusioni cui giungono dottrina e giurisprudenza con riguardo agli studi epidemiologici siano estensibili ai sistemi di IA di cui conosciamo il margine di errore e l'accuratezza. E infatti, le informazioni che il produttore ricava durante la fase di test del sistema sono equiparabili a quelle ottenute da un'indagine epidemiologica che ci dice che, in presenza di determinate circostanze, la probabilità di sviluppare una certa patologia è pari a una certa misura percentuale, così come i test di verifica di un sistema di IA ci rivelano la possibilità, misurata in percentuale, che quel sistema fornisca una risposta corretta a seguito del suo impiego. La conclusione che traiamo da ciò è che i dati forniti dal produttore non ci consentono da soli di stabilire se la risposta data dal singolo sistema in quel caso specifico sia corretta o errata, quindi se il verificarsi dell'evento penalmente rilevante sia riconducibile a un errore fatto dal sistema nel momento in cui veniva impiegato.

Non possiamo allora affermare che, poiché il sistema di riconoscimento di immagini inserito in un sistema di guida autonoma ha una percentuale di errore del 20%, allora il mancato riconoscimento di un ostacolo sulla carreggiata è sicuramente riconducibile a quel 20% di casi in cui il sistema si potrebbe sbagliare. Si tratterebbe di un ragionamento sviluppato sulla base di dati statistici, che nulla dicono sul caso concreto, così come l'epidemiologia non può svelare le cause della comparsa della malattia in quel singolo soggetto<sup>117</sup>; diversamente si darebbe rilevanza a probabilità calcolate in astratto per risolvere una dinamica eziologica concreta, con l'effetto di sostituire l'evento con il rischio che lo stesso si verifichi. Questo modo di affrontare l'accertamento del nesso eziologico darebbe perciò rilevanza al pericolo anche in reati che strutturalmente sono reati di danno e richiedono un evento. Basare l'accertamento causale esclusivamente sull'accuratezza del sistema determina il passaggio dalla "causalità" alla "mera

---

<sup>117</sup> Osservazioni simili sono sviluppate da V. Manes, seppur con riferimento all'uso di algoritmi nella decisione giudiziale: *Sotto un terzo profilo, analogamente correlato ai precedenti, emerge il problema della stessa fruibilità dell'algoritmo nel contesto del processo penale, e dei suoi peculiari criteri di giudizio, al di là degli ostacoli – forse non insuperabili – frapposti dal divieto di perizia criminologica stabilito dall'art. 220 c.p.p. e dall'uso degli algoritmi predittivi in chiave di valutazione della pericolosità. Ci si riferisce, in particolare, alle difficoltà – sul piano cognitivo e del facts-finding – di conciliare il criterio probatorio dell'"oltre ogni ragionevole dubbio" – la clausola BARD – con l'utilizzo di un software, in seno al quale solo apparentemente l'algoritmo fornisce maggior certezza. Basti pensare al rischio di affidare l'accertamento causale – ben oltre l'apporto che si vorrebbe trarre dall'epidemiologia – a correlazioni basate su dati statistici recepite e poste a base, appunto, degli algoritmi: dove giustamente si è evidenziato che "even with masses of data, there is no automatic technique for turning correlation into causation"*, V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *DisCrimen*, 15.5.2020, pp. 16-17.

causabilità”, avvicinando la causalità al pericolo posto che, mentre nella prima l’evento si è verificato, nel secondo esso è ritenuto solo possibile<sup>118</sup>. L’ancoraggio al paradigma nomologico sarebbe solo nominale, poiché si assisterebbe a un mutamento della fisionomia dell’evento. Una simile flessibilizzazione dei canoni di ascrizione della responsabilità non può essere praticata in un sistema penale che si basa sui principi di legalità e personalità della responsabilità solo per favorire la gestione giudiziaria delle incertezze e dei dubbi della post-modernità<sup>119</sup>. Il valore prognostico dei dati del test di validazione dei sistemi di IA può avere un ruolo nel momento processuale dedito all’accertamento del contributo causale che il sistema ha avuto nella determinazione dell’evento, ma non può esaurirlo.

Come si è in parte già anticipato, il dato statistico relativo alla percentuale di errore del sistema potrà avere rilevanza se corroborato da altre informazioni che consentano di escludere l’intervento di fattori causali alternativi, rappresentati, ad esempio, dal malfunzionamento delle componenti materiali del sistema, da un utilizzo improprio del manufatto da parte del compratore o da attacchi hacker provenienti dall’esterno che manomettono il funzionamento del sistema. Le cause allora possono essere diverse e appiattare l’accertamento del nesso di condizionamento sui soli dati statistici risultanti dal test di validazione significherebbe rinunciare al rigore che la teoria delle leggi scientifiche richiede per la causalità per adeguarla ai principi di legalità e di responsabilità personale di cui si è sopra detto. Solo nel momento in cui siano escluse cause diverse dall’errore algoritmico il dato del test di validazione può acquisire un valore probatorio centrale: se non vi sono altri fattori determinanti l’evento, ebbene deve ritenersi che esso è da ricondurre a un errore fatto dall’algoritmo, e ciò è comprovato dal margine di errore che emerge dal test di validazione, che offre l’immagine di un algoritmo che può sbagliare; e se non si accertano fattori causali alternativi allora la conclusione non può che essere che l’algoritmo ha sbagliato.

Si rende quindi necessaria la verifica volta ad escludere l’intervento di cause alternative all’errore algoritmico, verifica che non si colloca più sul piano della causalità generale, ma di quella individuale. Questa seconda fase di accertamento non darà sempre

---

<sup>118</sup> È questo il ragionamento proposto da O. Di Giovine in *Lo statuto epistemologico*, *op.cit.*, p. 658-659, ma come critica al modello nomologico di causa, dove le leggi scientifiche servirebbero alla rappresentazione di un risultato futuro ritenuto probabile o possibile, ma non consentirebbero di spiegare il singolo nesso causale.

<sup>119</sup> In questo senso, A. Gargani, *op. cit.*, p. 405.

luogo a un momento processuale semplice: l'opacità informatica che caratterizza soprattutto alcuni sistemi di IA complica la verifica a livello individuale del nesso eziologico e ciò alimenta la tentazione di ricadere su forme di flessibilizzazione dell'accertamento causale. È quindi venuto il momento di capire se esistono dei rimedi all'opacità informatica del singolo sistema di IA ovvero degli strumenti di cui il giudice possa servirsi per affrontare l'oscurità che avvolge il momento individuale della verifica causale.

#### **4. La causalità individuale o specifica**

Sin qui abbiamo dato atto di come la dottrina e la giurisprudenza concordino nel ritenere che l'accertamento giudiziale non possa essere affidato a un "sillogismo" che ricavi la conclusione sull'esistenza del nesso eziologico dal grado di evidenza della legge scientifica<sup>120</sup>. Come ha precisato l'intervento delle S.U. Franzese, il problema causale non va affrontato sul piano astratto, cioè dei modelli nomologici, poiché il giudice deve verificare in concreto la capacità esplicativa dell'enunciato, in particolare escludendo spiegazioni causali alternative dell'evento.

Lo schema metodologico del giudizio bifasico, quale modello euristico delineato dalle S.U. nel 2002, ritiene la causalità individuale provata a condizione che siano escluse spiegazioni alternative dell'evento, secondo il modello logico abduttivo, come poco sopra riportato: in questo modo sarebbe soddisfatta, anche attraverso generalizzazioni statistiche medio-basse, l'aspettativa di certezza che sorge dal principio della responsabilità personale e da quello di tassatività, il quale, abbracciando l'intera fattispecie, investe anche il nesso di condizionamento. L'esclusione di decorsi causali alternativi rappresenta un momento fondamentale nell'accertamento processuale: alla sua verifica è legato il raggiungimento della certezza processuale o "certezza pratica"<sup>121</sup> che consiste nell'alta credibilità razionale dell'ipotesi; una certezza normo-valutativa in quanto fondata su un ragionamento induttivo-argomentativo e non deduttivo<sup>122</sup>.

Il procedimento eliminativo di ipotesi alternative di spiegazione dell'evento era stato accolto anche in altri ordinamenti europei per l'accertamento della responsabilità penale

---

<sup>120</sup> C. Piemontese, *Il principio dell'"oltre il ragionevole dubbio", tra accertamento processuale e ricostruzione dei presupposti della responsabilità penale*, *Dir. pen. proc.*, 2004, 6, pp. 757 ss.

<sup>121</sup> M. Romano, *Il nesso causale*, *cit.*, p. 905.

<sup>122</sup> R. Bartoli, *Il problema della causalità penale*, *cit.*, p. 36.

da prodotto e in dottrina, accanto a chi lo ha criticato<sup>123</sup>, v'è chi ha da subito avvertito del rischio che l'accertamento di fattori alternativi si trasformasse in una formula utilizzata come clausola di stile per celare un vuoto nell'accertamento<sup>124</sup>. Si tratta di un timore fondato, perché se è vero che l'eliminazione di fattori alternativi (attraverso un procedimento per gradi che riesce a restringere il campo d'indagine a una sola ipotesi) riesce a chiarire che in un certo caso sia intervenuta una determinata ipotesi causale, è altresì vero che per riuscire in ciò, a rigore, dovrebbe essere richiesta la conoscenza di tutte le cause che portano alla verifica dell'evento<sup>125</sup>. Nondimeno, non sempre la conoscenza di tutti i fattori scatenanti l'evento è nota e la tentazione è quella di sostenere, seppur tacitamente, la tesi opposta, che ritiene sufficiente l'esclusione delle sole cause note, con la conseguenza che si farebbe gravare sull'imputato il peso dell'incertezza scientifica, in contrasto con la presunzione d'innocenza<sup>126</sup>, e che si consentirebbe l'ingresso di un aumento del rischio "processualmente mascherato"<sup>127</sup>.

Ma ancor prima di procedere alla prova per esclusione, è necessario che l'antecedente della cui efficacia causale si discute sia effettivamente inserito nell'insieme delle possibili cause dell'evento concreto entro le quali poi procedere con la prova per esclusione. Secondo la dottrina, l'inserimento di un antecedente tra le possibili cause di un evento richiede la dimostrazione dell'esistenza di casi in cui antecedenti del tipo di quello preso in considerazione abbiano provatamente causato un evento del tipo di quello verificatosi nel caso di specie, cioè che almeno in qualche occasione quell'antecedente abbia effettivamente cagionato un evento analogo a quello del caso concreto<sup>128</sup>. Si tratta di un passaggio che presenta già delle difficoltà, cui segue poi la prova per esclusione, che richiede, come si è visto, la conoscenza di tutti gli antecedenti eziologicamente in grado

---

<sup>123</sup> F. Stella, *Giustizia e modernità, cit.*, p. 359 ss., secondo il quale quando ricorrono leggi con frequenze medio-basse, il criterio dell'esclusione di ipotesi alternative riesce a selezionare la causa penalmente rilevante solo se vengono ricostruiti tutti gli anelli della catena causale; questa prova, che riuscirebbe ad azzerare il "peso" della pluralità delle cause, è però diabolica. Il criterio dell'esclusione di altre cause nelle ipotesi di leggi con frequenze medio-basse, allora, incorrerebbe in una fallacia, poiché dà per scontato che vi sia la prova che gli eventi cui si riferiscono le frequenze medio-basse siano causati da altri eventi, quando si tratta, invece, di una premessa che deve essere dimostrata. Per le critiche sviluppate dagli autori stranieri al procedimento per esclusione, si rinvia a M. Romano, *Nesso causale, cit.*, pp. 902-903, il quale invece si mostra a favore dell'accertamento del rapporto causale per via di esclusione.

<sup>124</sup> O. Di Giovine, *Probabilità, cit.*, pp. 2185-86.

<sup>125</sup> R. Bartoli, *op. cit.*, p. 36, p. 65, p. 83 ss.; Id., *Diritto penale e prova scientifica, cit.*, p. 100; D. Pulitanò, *Gestione del rischio, cit.*, p. 793; F. D'Alessandro, *Le frequenze medio-basse, cit.*, p. 4837; M. Donini, *Il garantismo della condicio, cit.*, p. 511.

<sup>126</sup> F. D'Alessandro, *op. cit.*, p. 4837.

<sup>127</sup> M. Romano, *op. cit.*, p. 906.

<sup>128</sup> F. D'Alessandro, *op. cit.*, p. 4834 e 4838.

di produrre l'evento, cioè di ogni alternativa "seria" e "plausibile" (e non tutte le alternative astrattamente considerate)<sup>129</sup>.

Per declinare più in concreto questo procedimento attinente alla fase processuale dell'accertamento del nesso eziologico, prendiamo l'esempio dell'insorgenza di patologie. In questi casi, per stabilire il nesso di condizionamento tra condotta ed evento, è necessario conoscere tutti i fattori scatenanti di quella malattia; tuttavia, mentre tale requisito può essere soddisfatto in patologie che hanno una sola causa, è più difficile che ciò avvenga allorché la patologia ha un'insorgenza multifattoriale e lo è ancora di più quando non tutte le cause scatenanti la malattia sono note<sup>130</sup>. In siffatte ipotesi, l'accertamento della derivazione causale della malattia dalla condotta dell'imputato non può essere accertata con la certezza richiesta nel processo penale, che obbedisce al principio dell'*oltre ogni ragionevole dubbio*. Fintantoché tutte le cause scatenanti di una certa patologia non saranno note, il dubbio ragionevole sulla sua eziologia esisterà.

Lasciando il contesto delle patologie e ritornando a quello dei danni da uso di sistemi intelligenti, potremmo forse ritenere che sarà più semplice individuare tutti gli antecedenti in grado di produrre l'evento del tipo di quello verificatosi. Per sostenere questo pensiero, partiamo da un esempio: un incidente in cui è coinvolta un'auto con sistema di guida autonoma potrà essere causato o da un problema del software ovvero dalla condotta umana, non essendo immaginabili altri antecedenti in grado di cagionare un evento analogo a quello osservato nel caso concreto, e questo faciliterà la successiva prova per esclusione delle altre possibili cause.

Nondimeno, la complessità dell'accertamento del nesso di derivazione permane sotto altri fronti. Difatti, a risultare più problematici saranno, probabilmente, la ricostruzione degli anelli causali intermedi, da un lato, e l'individuazione del contributo effettivamente rilevante da parte degli antecedenti che hanno preceduto la realizzazione dell'evento, dall'altro. Considerando proprio quest'ultimo aspetto, si può immaginare che difficilmente ci si troverà a dover provare un enunciato causale in cui tra azione *x* e evento *y* vi sia un rapporto diretto, del tipo "x ha causato y"; piuttosto potrà essere frequente la presenza di più concause, secondo l'enunciato "x1, x2 ... hanno causato y"<sup>131</sup>. Questo scenario mi sembra assai realistico considerato che le indicazioni rivolte dal legislatore

---

<sup>129</sup> M. Romano, *op. cit.*, p. 909.

<sup>130</sup> R. Bartoli, *Il problema della causalità penale, cit.*, p. 65.

<sup>131</sup> Sull'enunciato relativo al nesso causale, si rinvia a M. Taruffo, *op. cit.*, p. 89 ss.

europeo agli Stati membri delineano sempre un utilizzo dei sistemi di IA supervisionato dall'uomo, quindi è probabile che alla realizzazione di un evento di danno contribuisca anche l'uomo, cioè l'utente, oltre che il sistema.

Il problema della ricostruzione degli anelli causali intermedi, invece, si pone poiché, a meno di aderire alla tesi della responsabilità del sistema di IA, il contributo causale del sistema andrà collegato alla figura del produttore e questo solleva ulteriori riflessioni.

I sistemi di IA ci appaiono come pezzi monolitici, ma in realtà non è raro che essi costituiscano l'insieme di attività di più soggetti (più persone fisiche all'interno di uno stesso ente o anche più enti), che si impegnano nelle fasi di progettazione, fabbricazione e assemblaggio. Questo insieme di energie produttive lascia ipotizzare la presenza di vari passaggi causali, secondo l'enunciato "x ha causato x1, che ha causato x2...che ha causato y". Si tratta, allora, di ricostruire la catena causale e di individuare gli errori che sono stati fatti nella fase produttiva.

Un simile accertamento dovrà essere effettuato sia quando emerge una responsabilità dell'utilizzatore, al fine di escludere una responsabilità del produttore, che quando all'utente nessun rimprovero potrà essere mosso: infatti, appurato che la causa dell'evento è dovuta a un malfunzionamento del sistema, sarà necessario isolare le responsabilità all'interno della filiera produttiva, cercando di comprendere quali soggetti inquadrabili nella più generale categoria di "produttore" siano responsabili.

Al di là del caso in cui l'evento sia condizionalmente derivato dalla sola condotta dell'utilizzatore, per il resto bisognerà sempre analizzare il funzionamento del sistema intelligente, per capire *dove* si colloca l'errore, cioè in quale momento del funzionamento il sistema non ha risposto come avrebbe dovuto, così da individuare il soggetto responsabile all'interno della catena produttiva.

Si tratta di una verifica non semplice. Abbiamo visto come sia articolata la ricostruzione del processo decisionale da parte dell'intelligenza artificiale e come essa sia particolarmente complessa nel caso di sistemi che abbiamo definito di tipo "biologico", in quanto riproducenti la struttura neuronale del cervello per l'elaborazione della decisione.

La dottrina penalistica non ha mancato di sottolineare questo aspetto, tanto è vero che in molti avvertono del rischio di trovarsi dinanzi a una vera e propria *black box*: una scatola nera all'interno della quale è impossibile leggere le dinamiche causali<sup>132</sup>.

Arrestando qui il discorso, saremmo portati a ritenere che la diffusione dell'utilizzo di tecnologie di IA per l'espletamento di attività quotidiane lasci presagire un aumento della fenomenologia di danni di cui sia poco chiaro il meccanismo di produzione. Questa affermazione ritengo sia vera, ma penso che abbia anche bisogno di alcune specificazioni, che di seguito tenterò di sviluppare.

Prima, però, bisogna brevemente ricordare che il problema dell'opacità si pone nel momento in cui l'uomo si avvicina alla macchina intelligente non per utilizzarla, ma per "comprenderla". Questa "comprensione" conosce declinazioni diverse, dovendosi distinguere tra esplicabilità, trasparenza e interpretabilità. La prima concerne la spiegazione del processo decisionale detta *explanation*, che è l'informazione o 'spiegazione' fornita dal sistema in relazione alle cause dell'output. La trasparenza attiene all'informazione relativa al funzionamento strutturale della macchina, cioè alle sue componenti, e ai dati utilizzati per l'addestramento. Le informazioni ottenute da sistemi esplicabili e trasparenti devono poi essere "tradotte" in un linguaggio che è comprensibile per l'uomo e per questo si fa riferimento all'interpretabilità, da ricondurre all'atto di *interpretation* fatto dall'agente; l'interpretazione, pertanto, coincide con la "lettura" del procedimento decisionale da parte dell'agente, partendo dalle 'spiegazioni' fornite dal sistema<sup>133</sup>. Pertanto, il discorso sul contributo eziologico che in concreto un algoritmo apporta alla causazione di un evento penalmente rilevante va condotto notando che il giudice in sede processuale dovrà interpretare il funzionamento dell'algoritmo ricorrendo a tecniche che spieghino lo stesso e che tale accertamento sarà tanto più semplice quanto più trasparente sia il sistema di IA coinvolto.

---

<sup>132</sup> R. Borsari, *Intelligenza artificiale e responsabilità penale; prime considerazioni*, *Media Laws*, p. 265; S. Beck, *Google cars, software agents, autonomous weapons systems – New challenges for criminal law?*, in E. Hilgendorf, U. Seidel (eds.), *Robotics, Autonomics, and the Law*, Baden-Baden, 2017, 227 ss.; A. Cappellini, *Machina delinquere potest?*, *DisCrimen*, 27.3.2019, p. 7; S. Quattrocolo, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva"*, in *Cass. pen.*, 2019, p. 1759; V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *DisCrimen*, 15.5.2020, p. 14.

<sup>133</sup> R. Tomsett, D. Braines, D. Harborne, A. Preece, S. Chakraborty, *Interpretable to whom? A role based model for analyzing interpretable machine learning systems*, 2018, *arXiv.org*, p. 9.

#### 4.1 L'incertezza: la causalità individuale rinchiusa in una scatola nera. Opacità *ex post* e *explainable IA*

Non è la prima volta che il diritto penale si trova ad affrontare eziologie difficili o impossibili da ricostruire. Le scatole nere, allora, sono tante e quello dei sistemi di IA è solo un altro caso che si aggiunge alla già lunga lista delle sfide poste dalla modernità<sup>134</sup>.

Si pensi, ad esempio, alle malattie professionali, il cui decorso patogeno ha già investito i giudici del problema della ricostruzione di decorsi causali in cui alcuni passaggi sono rimasti oscuri.

L'oscurità causale ha dominato anche il contesto processuale dei danni da prodotto, per i quali la dottrina già adoperava l'espressione *black box* causale<sup>135</sup>. Tra i casi più noti affrontati dalla giurisprudenza penale di alcuni Stati europei, basti qui ricordare i processi *Contergan*, sul preparato farmaceutico talidomide, il caso *Lederspray*, sullo spray per pellami, e il caso dell'*Holzschutzmittelfall*, prodotto per la protezione del legno, tutti decisi dalla giurisprudenza tedesca; ancora il caso dell'olio di colza che ha occupato la giurisprudenza spagnola; il caso francese *Benoliel*, sull'uso dell'insetticida *Néocide Fulgurant*<sup>136</sup>.

La dottrina che ha commentato tali decisioni aveva già sottolineato la crisi cui era andata incontro la categoria penale della causalità, ritenuta integrata sulla base di stime incerte, secondo logiche tipiche della responsabilità civile “per rischio”, (in base alla quale il produttore deve sopportare il rischio della propria attività produttiva, che non deve quindi essere scaricato sulle vittime), con conseguente abbandono del modello nomologico e frustrazione dei principi di determinatezza e frammentarietà dell'illecito penale<sup>137</sup>.

Oggi ci riferiamo all'opacità causale nella responsabilità da prodotto “intelligente”, in cui la vera differenza con i casi sopra menzionati è che quest'ultima non è ancora stata oggetto di trattazione nei tribunali italiani. L'assenza di decisioni giurisprudenziali ne rende in questa sede diversa la trattazione, poiché non si può che ragionare in astratto e per ipotesi.

---

<sup>134</sup> Una sintesi efficace sulle tendenze della prassi dinanzi ai nuovi rischi e le risposte della dottrina, G. De Francesco, *L'imputazione del reato e i tormenti del penalista*, in *Scritti per Federico Stella*, cit., pp. 513 ss.

<sup>135</sup> C. Piergallini, *op. cit.*, p. 13.

<sup>136</sup> Per l'analisi delle decisioni si rinvia a C. Piergallini, *op. cit.*, pp. 50 ss.

<sup>137</sup> *Ibidem*, Parte I, cap. II, pp. 140 ss.

Sul tema dell'opacità causale, va subito notato che la questione è sicuramente da circoscrivere ad alcuni sistemi di IA, in particolare a quelli che ricorrono a sistemi di *machine learning*, quindi di autoapprendimento. Tra questi, come si è visto nel primo capitolo, i sistemi di tipo “logico”, come gli alberi decisionali, presentano meno problemi, poiché adottano un modello di apprendimento *top-down*, che consente di ricostruire i passaggi decisionali del sistema. Va però detto che anche questi sistemi diventano opachi nel momento in cui assumono dimensioni notevoli, mentre sono chiari nella loro struttura logico-evolutiva solo se hanno un'estensione contenuta. Un altro elemento che offusca la dinamica decisionale, oltre alle dimensioni notevoli che gli alberi decisionali possono assumere, è l'importante lavoro svolto dai programmatori sui dati, che non vengono proposti al sistema in modo grezzo, ma vengono rielaborati per renderli comprensibili sulla base delle regole logiche che guidano questo modello di IA<sup>138</sup>.

L'opacità è presente in modo ancor più rilevante nei modelli *bottom-up*, che non seguono un processo logico di apprendimento, in particolare nelle reti neurali caratterizzate da apprendimento profondo (*deep learning*). In questo caso l'opacità è connessa non alle dimensioni raggiunte dal modello (come avviene per gli alberi decisionali), ma al modo in cui i sistemi di *deep learning* sono programmati, poiché i programmatori non stabiliscono come un determinato compito è assolto dal sistema, ma solo le condizioni in presenza delle quali una soluzione può essere proposta dal sistema<sup>139</sup>, che apprende dall'esperienza. Ne derivano non poche difficoltà nel prevedere come l'algoritmo “si comporterà” durante il suo utilizzo. Questa imprevedibilità *ex ante* è il frutto della modalità di funzionamento dei sistemi di *machine learning* che però non rileva ancora sul piano dell'accertamento del nesso di condizionamento. Essa, come si è già visto *supra*, SEZIONE I di questo capitolo, dovrà essere analizzata in sede di verifica della colpa, cioè nel corso del giudizio di prevedibilità e evitabilità dell'evento.

Il piano dell'accertamento causale è invece interessato alla comprensione di come il sistema di IA ha funzionato nel caso concreto. Anche questo non è un procedimento lineare e di facile risoluzione, che coinvolge i sistemi di *machine learning*, che è alla base della opacità che li caratterizza *ex post*, cioè una volta che la decisione è stata assunta. È

---

<sup>138</sup> Z. C. Lipton, *The mythos of model interpretability*, *ArXiv*, 2016, 1606, par. 3.1 e 4.1; K. A. Creel, *Function and User-Satisfaction in Explainable AI*, in *AISB 2021 Symposium Proceedings: Overcoming Opacity in Machine Learning* (ed. C. Zednik, . Boelsen), p. 3.

<sup>139</sup> C. Zednik, *Solving the Black Box Problem: a normative framework for explainable AI*, *Phil.Tech.*, 2021, p. 267.

bene quindi circoscrivere l'opacità causale a questo momento dell'interpretazione del funzionamento dell'algoritmo, distinguendo e isolando questa problematica da quella dell'imprevedibilità *ex ante* che nella sistematica del reato involgerà l'elemento della diligenza del reato colposo.

Il giudice, nel momento in cui si occupa dell'accertamento causale, sarà chiamato a interpretare il processo decisionale dell'algoritmo, che è un atto *a posteriori*, cioè che segue la soluzione data e che solo in parte coincide con le fasi iniziali di produzione del *software*.

Per fare questo il giudice può avvalersi di tecniche di interpretazione dei modelli opachi che intervengano "*after-the-fact*", che una parte della letteratura scientifica raggruppa sotto l'espressione "*post-hoc interpretations*"<sup>140</sup>. Accanto alle tecniche di *machine learning* hanno conosciuto sviluppo anche studi che cercano di fare luce su questa oscurità che intercorre tra i dati di input e l'output, elaborando sistemi di spiegazione, raggruppati sotto la denominazione *explainable AI (XAI)*<sup>141</sup>. Possiamo immaginare che l'interesse per questi studi non proviene solo dal modo del diritto, ma anche da quello produttivo, posto che sistemi imperscrutabili risultano non affidabili e quindi non allettanti per gli investitori o per gli utenti consumatori<sup>142</sup>.

L'opacità del sistema si caratterizza per due elementi: a) la presenza di un agente, che utilizza il sistema o che si confronta con esso; b) la non conoscenza degli elementi del sistema ritenuti epistemicamente rilevanti (*epistemically relevant elements*)<sup>143</sup>.

---

<sup>140</sup> Z. C. Lipton, *op. cit.*, par. 3.2, terminologia ripresa anche da R. Tomsett, D. Braines, D. Harborne, A. Preece, S. Chakraborty, *Interpretable to whom?*, *cit.*, p. 9, e nel lavoro di G. Montavon, W. Samek, K.R. Müller, *Methods for interpreting and understanding deep neural networks*, *Digital signal processing*, 73, 2018, pp. 1-15.

<sup>141</sup> J. Burrell, *How the machine 'thinks': Understanding opacity in machine learning algorithms*. *Big Data & Society*, 3(1), 2016; Z. C. Lipton, *The mythos of model interpretability*, *ArXiv*, 2016, 1606; C. Zednik, *op. cit.*

<sup>142</sup> La non verificabilità del processo decisionale degli algoritmi penalizza fortemente il loro uso in settori strategici, come quello medico-diagnostico, dove si fatica a ritenere affidabili le diagnosi dei software. Un caso oggetto di studio è quello del sistema denominato *Deep Patient*, rete neurale nella quale erano stati inseriti dati di 700 mila pazienti raccolti in dodici anni. *Deep Patient* riusciva a predire la comparsa di future patologie, tra cui diabete, schizofrenia e tumori, ma la sua opacità non consentiva di spiegare come la diagnosi era svolta e per questo i medici non erano in grado di "apprendere" dall'algoritmo, Miotto, R. *et al.* *Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records*, *Sci. Rep.* 6, 26094, (2016); W. J. von Eschenbach, *Transparency and the Black Box Problem: Why We Do Not Trust AI*, *Phil. and Tech.*, 1. Sept. 2021, par.2. Sull'importanza della trasparenza e interpretabilità dell'algoritmo in ambito medico, si rinvia a J. C. Bjerring, J. Busch, *Artificial Intelligence and Patient-Centered Decision-Making*, *Phil. and Tech.*, 8 jan. 2020, pp. 349-371; M. Hutson, *Self-taught Artificial Intelligence Beats Doctors at Predicting Heart Attacks*, *Science*, 2017.

<sup>143</sup> P. Humphreys, *The philosophical novelty of computer simulation methods*, *Synthese*, 2009, 169(3), pp. 615-626: "Here a process is epistemically opaque relative to a cognitive agent *X* at time *t* just in case *X* does not know at *t* all of the epistemically relevant elements of the process. A process is essentially

Pertanto, il compito della *explainable AI* è rendere noti gli elementi epistemicamente rilevanti dell'IA<sup>144</sup>, per spiegare *come* il sistema è giunto a una certa soluzione, *dove* si colloca l'eventuale errore, *perché* il sistema ha dato un certo risultato o *perché* il sistema ha errato<sup>145</sup>.

È da condividere l'idea che non è sempre necessario rispondere a tutti questi quesiti, poiché la risposta anche a uno o a due di essi potrebbe già essere sufficiente e ciò dipende da *chi* pone la domanda<sup>146</sup> e in quale contesto è svolta l'analisi del sistema di IA. Un aspetto, questo, che non è di scarso rilievo, poiché va considerato che le tecnologie di spiegazione non sempre riescono a risolvere tutti gli aspetti dell'opacità. A questi limiti è possibile far fronte utilizzando la tecnologia appropriata al tipo di *stakeholder*, cioè al soggetto che si interroga sul funzionamento del sistema. Ad esempio, pensando alla posizione dell'utilizzatore, questi sarà interessato a capire *come* il sistema utilizza i dati inseriti e *perché* viene dato un certo output, mentre un tecnico che lavora alla produzione del *software* avrà interesse a comprendere *dove* si colloca l'errore e *perché* si è verificato. In alcuni contesti a queste informazioni potranno essere del tutto disinteressati gli utenti che ottengano il risultato sperato (è questo il caso delle aziende che utilizzano i *cookies* pubblicitari, che non hanno interesse a capire perché un certo annuncio è stato offerto a una determinata persona se comunque avrà un incremento delle vendite<sup>147</sup>).

Di certo, però, il discorso cambia per il contesto di nostro interesse, che è quello processuale, in cui a interagire con il sistema di IA è il giudice e, prima ancora, la pubblica accusa e la difesa e dove vi è in gioco la libertà dell'imputato. Ne deriva che in sede processuale i problemi esplicativi si accentuano, poiché non ci si potrà accontentare di spiegazioni parziali e, inoltre, sarebbe necessario ricostruire lo storico degli output e testare il sistema facendolo interagire con nuovi dati<sup>148</sup>. Si tratterà di una perizia che spesso potrà risultare molto dispendiosa. Il giudice, infatti, dovrebbe essere in grado di rispondere alle seguenti domande:

1. qual è il compito assegnato al sistema (*what*).

---

*epistemically opaque to X if and only if it is impossible, given the nature of X, for X to know all of the epistemically relevant elements of the process*", p. 618.

<sup>144</sup> C. Zednik, *op. cit.*, p. 269.

<sup>145</sup> C. Zednik, *op. cit.*, pp. 272 ss.

<sup>146</sup> R. Tomsett, D. Braines, D. Harborne, A. Preece, S. Chakraborty, *Interpretable to whom?*, *cit.*

<sup>147</sup> J. A. Kroll, *cit.*, p. 3.

<sup>148</sup> R. Tomsett, et al., *cit.*, p. 13.

2. come funziona il sistema (*how*). Questo consentirà di comprendere se si è dinanzi a un sistema di IA che impiega o meno la tecnica di ML e se sì, quale.

3. perché il sistema è giunto a formulare quella decisione (*why*). Si tratta di analizzare non solo il funzionamento, ma anche i dati utilizzati per l'addestramento e i dati di cui il sistema si è "nutrito" prima di prendere la decisione.

4. quali sono le componenti del sistema hardware, inteso anche nella sua parte robotica (quindi fisica), che vengono in gioco nel caso di specie (*where*). In questo modo potranno essere isolate le responsabilità all'interno della catena produttiva.

#### **4.2 Verso il superamento dell'opacità causale?**

Il ricorso ai modelli di spiegazione può consentire al giudice di rispondere ai quesiti nel momento dell'accertamento del nesso causale. Immaginiamo allora un caso in cui l'impiego di tali tecniche può agevolare questo passaggio nella decisione giudiziale.

Una vettura dotata del sistema di pilota automatico non si arresta in prossimità di un incrocio provocando delle lesioni al passeggero di un'altra vettura che non aveva l'obbligo di arrestarsi in quanto godeva del diritto di precedenza. Il conducente dell'auto che non si arrestava all'incrocio dichiarava che il sistema non aveva inviato alcun segnale di ripresa del controllo della vettura e per questo non riusciva a intervenire per arrestare la vettura, con conseguente verifica del sinistro.

Il giudice dovrà accertare il contributo causale del conducente, da un canto, e del sistema, quindi del produttore, dall'altro.

Poniamo adesso l'attenzione sulla ricostruzione del processo che ha condotto l'algoritmo dal dato di *input* (immagine dell'incrocio) all'*output* (mancato riconoscimento del segnale di dare precedenza e mancato arresto), tralasciando per il momento la verifica del contributo causale del conducente.

Poniamo che la perizia accerti che a non aver funzionato bene è il *software* deputato alla visualizzazione delle immagini, che non riconosceva l'incrocio e ometteva di inviare i segnali al sistema centrale. Si pone evidentemente un problema di mancato riconoscimento degli elementi dell'immagine e per ricostruire i passaggi che intercorrono tra l'*input* e l'*output* il giudice potrebbe ricorrere ai modelli di *post-hoc explanation o interpretation* deputati alla lettura dei sistemi di classificazione di immagini. Tra questi è possibile annoverare la *input heatmapping*, che consente di individuare gli elementi dei dati di *input* che hanno un peso rilevante nella formazione dell'*output*, così da creare delle

“mappe di calore”<sup>149</sup>. Nel caso in cui siano applicati a modelli di classificazione delle immagini, la tecnica della mappa di calore consente di individuare i *pixel* che il sistema prende in considerazione tra quelli che compongono l’immagine di *input*, così spiegando *come e perché* si è giunti a quel determinato *output* e isolando gli elementi grafici ai quali l’algoritmo dà un peso maggiore. L’algoritmo sarà “difettoso” se dà un peso maggiore a ciò che appare nello sfondo piuttosto che al soggetto principale che deve essere classificato<sup>150</sup>.

Applicando questa tecnica ai sistemi di guida autonoma, sarà possibile comprendere quali elementi di un’immagine il sistema di IA considera più rilevanti all’atto di individuare un incrocio o una vettura o un pedone. Così nel caso oggetto dell’esempio, l’incidente dovuto al mancato arresto all’incrocio potrà essere dovuto all’incapacità del sistema di distinguere l’immagine dell’incrocio poiché valorizza solo alcuni elementi visivi (ad esempio l’assenza di segnaletica orizzontale o verticale) e non altri che contribuiscono a delineare in modo chiaro la presenza dell’intersezione stradale (ad esempio l’interruzione del *guard rail*).

Questa tecnica consente allora al giudice di “spacchettare” l’algoritmo, che nasce opaco, e che è spiegato da un secondo algoritmo che illustra il processo decisionale seguito e il giudice vi farà ricorso in sede processuale così come già oggi si effettuano accertamenti autoptici o di laboratorio a posteriori.

Va altresì notato che per ricostruire la dinamica causale il funzionamento del sistema dovrebbe essere registrato e salvato, in modo che possa poi essere analizzato svolgimento concreto dei fatti. L’accesso ai dati in sede giudiziale sarà fondamentale poiché consentirà di lavorare sullo sviluppo dei fatti così come concretamente verificatisi, piuttosto che lavorare su mere ipotesi causali. Per ritornare all’esempio del sinistro stradale, l’accesso alla registrazione renderà conoscibili i dati captati dai sensori prima e durante lo scontro, i quali potranno anche essere comparati con i dati captati da altri veicoli eventualmente coinvolti nello scontro o anche non coinvolti ma presenti al momento del sinistro.

È allora necessario prevedere a livello legislativo la predisposizione della registrazione delle immagini, in modo simile a quanto accade già oggi con la mobilità aerea, e consentirne l’accesso all’autorità giudiziaria, sempre bilanciando questa necessità

---

<sup>149</sup> G. Montavon, *et al.*, *op. cit.*, pp. 4-6.

<sup>150</sup> È quanto accaduto al sistema di classificazione deputato a distinguere le immagini di cane husky da quelle di lupi, in cui la classificazione si concentrava sullo sfondo (presenza di neve per i cani husky) piuttosto che sulle caratteristiche dell’animale. C. Zednik, *op. cit.*, p. 275 ss.

con i diritti fondamentali, in particolare col diritto alla tutela della vita privata. La registrazione è attualmente prevista all'art. 12 della proposta di regolamento dell'aprile 2021 per i sistemi di IA ad alto rischio durante il loro funzionamento. In siffatto contesto, i produttori e gli utilizzatori dovrebbero essere tenuti a cooperare con l'autorità giudiziaria per fornire le informazioni utili<sup>151</sup>. L'accesso alla registrazione da parte delle autorità investigative è stata prevista dal legislatore francese all'art. 123-3 del codice della strada con la riforma dell'aprile 2021, che ha inserito delle disposizioni *ad hoc* sulla responsabilità del conducente e del produttore in caso di reati integrati all'atto della guida in caso di vetture dotate di sistemi di guida autonoma<sup>152</sup>, che saranno analizzate più avanti (CAPITOLO IV).

Il requisito dell'accessibilità ai dati che riportano il funzionamento consente di abbattere un ulteriore momento dell'opacità che potremmo definire *opacità fenomenologica concreta*. Il reale svolgimento dei fatti potrà essere ricostruito con più facilità poiché la dinamica dell'accaduto sarà conoscibile dal giudice grazie all'accesso ai dati registrati dal sistema.

---

<sup>151</sup> L'idea è già avanzata nella Risoluzione del Parlamento europeo dell'ottobre 2020, sopra analizzata, seppur con riferimento alla responsabilità civile. Si legge, in proposito al par. (19) che “*Per consentire all'operatore di dimostrare l'assenza di colpa o per consentire alla persona interessata di dimostrarne l'esistenza, i produttori dovrebbero essere tenuti a cooperare con entrambe le parti interessate, anche fornendo informazioni ben documentate*”; all'art. 10 della proposta di regolamento ivi contenuta, dedicato al concorso di colpa, al c. 2 si legge: “*Un operatore ritenuto responsabile può utilizzare i dati generati dal sistema di IA per provare il concorso di colpa della persona interessata, in conformità del regolamento (UE) 2016/679 e di altre leggi pertinenti in materia di protezione dei dati. La persona interessata può utilizzare tali dati anche come prova o ai fini di un chiarimento nell'ambito dell'azione per responsabilità*”.

<sup>152</sup> Art. 123-3 c. de la route : I.- *Sans préjudice des dispositions des articles 60-1,60-2,77-1-1 et 99-3 du code de procédure pénale, ont accès aux données du dispositif d'enregistrement des données d'état de délégation de conduite :*

1° *Les fonctionnaires du corps de commandement ou d'encadrement de la police nationale mentionnés aux articles L. 130-1 et L. 130-3 du présent code, lorsque le véhicule est impliqué dans un accident de la circulation ayant occasionné un dommage corporel ;*

2° *Les agents compétents pour constater les contraventions au présent code en application de l'article L. 130-4, à l'occasion des contrôles des véhicules et de leurs conducteurs ;*

3° *Le titulaire du certificat d'immatriculation du véhicule ou les personnes visées aux trois derniers alinéas de l'article L. 121-3, en cas de constatation d'une des contraventions mentionnées à cet article.*

II.-*Pour les fins précisées au I, le constructeur du véhicule ou son mandataire, au sens de l'article 3 du règlement (UE) 2018/858 du Parlement européen et du Conseil du 30 mai 2018, garantit l'intégrité des données mentionnées au premier alinéa ainsi que leur accès.*

*Dans le cas où le constructeur du véhicule ou son mandataire, au sens de l'article 3 du règlement (UE) 2018/858 du Parlement européen et du Conseil du 30 mai 2018, a accès à ces données à distance, lorsque le véhicule est équipé de moyens de communication permettant de les échanger avec l'extérieur de celui-ci, les modalités de cet accès et de conservation des données, dont la durée ne peut dépasser six ans à compter de la date de l'accident dans le cas prévu au 1° du I, ou un an à compter de la date des faits dans les autres cas, sont précisées par décret en Conseil d'Etat.*

Si può allora immaginare che in molti casi in cui sia coinvolto un sistema di IA nella produzione di eventi di danno penalmente rilevanti alcuni aspetti del decorso causale saranno ricostruibili con maggiore facilità proprio grazie alla conservazione dei dati di monitoraggio del funzionamento del sistema.

Altre tecniche di *post-hoc explanation* sono la *local rule-based explanations (LORE)* e la *local interpretable model-agnostic explanation (LIME)*.

LIME è un algoritmo che spiega le previsioni fatte da un sistema che si occupa di classificazione, illustrandone la decisione attraverso lo sviluppo di un modello interpretabile<sup>153</sup>. Ad esempio, se un sistema di diagnosi medica riconosce in un paziente il virus dell'influenza, senza spiegare le ragioni sottese a tale diagnosi, l'algoritmo LIME individua i dati che hanno determinato quella decisione; nella specie, indicherà nell'insieme dei dati di input i sintomi che hanno permesso l'individuazione del virus influenzale (ad esempio i sintomi 'difficoltà a respirare' e 'mal di testa') e quelli ritenuti irrilevanti per la classificazione (ad esempio il sintomo 'no fatica')<sup>154</sup>.

Le tecniche di *explainable IA*, alcune delle quali sono state appena illustrate, ci consentono di limitare la portata dell'obiezione basata sul *black-box* causale, per sostenere, al contrario, che è possibile l'impiego di *white-box testing regimes*. Sarebbe però affrettato considerare superata la questione dell'opacità.

L'*explainable IA* è un settore che è ancora in pieno sviluppo<sup>155</sup>, come del resto lo è quello dei modelli di *machine learning* su cui interviene, e le interpretazioni offerte possono essere imprecise o soggette ad errori. Accanto a chi esalta queste tecniche, vi è anche chi le analizza in modo critico, osservando che la spiegazione può essere imprecisa e non totalmente fedele al modello originario<sup>156</sup>.

Il rischio, allora, è che la spiegazione fornita sia una rappresentazione non accurata del modello originario, con evidente incompatibilità con le esigenze di certezza dell'accertamento del processo penale. Una spiegazione che rappresenti in modo impreciso il modello oggetto di accertamento è del tutto inutile e anzi, è fortemente

---

<sup>153</sup> M. T. Ribeiro, S. Singh, C. Guestrin, "Why Should I Trust You?": Explaining the Predictions of Any Classifier, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, pp. 1135–1144. ACM, 2016.

<sup>154</sup> *Ibidem*.

<sup>155</sup> J. A. Kroll, *The fallacy of inscrutability*, *Phil. Trans. R. Soc. A* 376, p. 11.

<sup>156</sup> C. Rudin, *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*, in *Nature Machine Intelligence*, 1(5), 206–215, May 2019.

dannosa in sede processuale: qui, dove regna la regola dell' *oltre ogni ragionevole dubbio*, occorrerà una *fedele rappresentazione del modello*.

Un altro aspetto da non sottovalutare è che la *post-hoc explanation* non offre una valutazione circa la correttezza o meno della decisione del modello. Questa valutazione è lasciata all'uomo, mentre l'algoritmo esplicativo riproduce semplicemente il processo decisionale rimasto ignoto, "spacchettando" o "aprendo" la scatola nera. Il rischio è che la distinzione tra valutazione della correttezza della decisione e ricostruzione del processo che porta alla decisione potrebbe divenire labile fino a scomparire, cosicché anche le risposte "errate" della macchina riceverebbero una loro spiegazione che potrebbe far apparire le stesse come "corrette" o "ragionevoli"<sup>157</sup>. La tematica è di grande rilievo se si pensa che la *explainable IA* si presta a divenire uno strumento che potrebbe confermare scelte errate o discriminatorie fatte dall'algoritmo, fomentando così i timori verso una società sottoposta alla "dittatura dell'algoritmo", in cui nuovi algoritmi confermano decisioni di altri algoritmi, anche se irragionevoli.

Simili risvolti patologici di sistemi nati come soluzione al problema dell'opacità, ma che nella sostanza ne originano dei nuovi, possono essere arginati in vario modo.

Anzitutto bisogna relativizzare il ruolo del modello di spiegazione: in quanto rappresentazione del modello originario, esso non può essere del tutto sovrapponibile al primo. Anche le tecniche di *XAI*, in quanto rappresentazioni, sono asimmetriche. Questo concetto è spiegabile attraverso degli esempi: quando affermiamo che un certo quadro rappresenta un certo personaggio, ad esempio Napoleone, sappiamo bene che non varrà il contrario, cioè che Napoleone non rappresenta il quadro; l'asimmetria non fa però venire meno il valore epistemico del quadro, che è un mezzo che facilita la conoscenza del soggetto rappresentandolo. Ancora, una mappa della metropolitana non è identica alla rete reale quanto a dimensione, tuttavia ci permette di conoscere la connessione tra le varie linee<sup>158</sup>. Del pari, un modello di spiegazione rappresenta le cause sottese all'*output* fornito dal modello originario e, sebbene non assomigli del tutto a quest'ultimo, ci consente comunque di acquisire una conoscenza del processo causale attraverso la sua rappresentazione, nonostante i due modelli non siano identici.

---

<sup>157</sup> J. A. Kroll, *op. cit.*, p. 10.

<sup>158</sup> L. Chan, *Explainable AI as Epistemic Representation*, in *AISB 2021 Symposium Proceedings*, cit., p. 8. L'esempio della rete metropolitana è ivi ripreso da R. Carnap, *The Logical Structure of the World: Pseudoproblems in Philosophy*, University of California Press, 1967.

L'attendibilità della *post-hoc interpretation* dovrà essere valutata dall'utente, o dal giudice in sede processuale. Questa valutazione dovrebbe tenere conto di alcune caratteristiche del sistema, individuando gli elementi del modello originario oggetto di rappresentazione; dovranno essere indicate le ragioni per le quali si è scelto di rappresentare solo certi elementi e non altri; la rappresentazione dovrà consentire di trarre inferenze affidabili sulle caratteristiche individuate<sup>159</sup>.

L'approccio verso le tecniche di spiegazione diviene così attivo e non passivo: il responso di questi modelli non potrà essere accolto in modo acritico dal giudice, il quale dovrà sempre valutare l'attendibilità di tali sistemi ricorrendo al suo prudente apprezzamento, tenendo a mente che si tratta di strumenti che rappresentano il modello originario.

Le criticità presentate dai metodi di *post-hoc explanation* spingono però verso la ricerca di soluzioni alternative, volte a incentivare (imporre?) la produzione di sistemi di *machine learning* che nascono già come interpretabili, in modo da superare l'opacità a monte, piuttosto che ricorrere a rimedi a valle<sup>160</sup>. Ciò consentirebbe di limitare il ricorso alle tecniche di *extent explanation*. Questo requisito andrebbe posto a livello legislativo, come precisato dagli interventi di *soft law* europei, che indicano l'esplicabilità tra i requisiti di una IA affidabile. Il legislatore, consapevole delle difficoltà riscontrabili in sede probatoria, dovrebbe regolare la produzione dei manufatti intelligenti in modo da favorire l'analisi e la verifica della decisione automatica. A tal proposito, ad esempio, potrebbe essere previsto un processo di decisione algoritmica "frazionato", stabilendo il raggiungimento di obiettivi intermedi come requisito per la prosecuzione nella decisione che, se non realizzati, sbarrano la strada al funzionamento dell'algoritmo. Si tratterebbe di una decisione "guidata" che permetterebbe di individuare il momento in cui l'algoritmo ha errato o non è stato in grado di proseguire nel suo compito<sup>161</sup>. In quel momento l'algoritmo dovrebbe smettere di funzionare e inviare un segnale all'utente, avvertendolo del mancato rispetto degli standard decisionali impostati dal produttore e inviando un messaggio di richiesta di ripresa del controllo da parte dell'uomo.

---

<sup>159</sup> *Ibidem*, p. 8.

<sup>160</sup> C. Zednik, H. Boelsen, *Preface: Overcoming Opacity in Machine Learning*, in *AISB 2021 Symposium Proceedings*, cit., p. 1.

<sup>161</sup> L'idea è sviluppata a partire dalle riflessioni di J.A. Kroll: "Once a system's requirements are established, it becomes necessary to digest them into a design and to develop an explicit set of goals or invariants for that design to uphold", *op. cit.*, p. 11.

Si può ancora osservare che le tecniche di *explanation* potrebbero essere superflue in caso di errore dovuto alla selezione dei dati iniziali. In tale evenienza, piuttosto che applicare tecnologie esplicative che analizzino in modo particolareggiato il funzionamento meccanico del sistema, va condotta un'analisi sui dati. Il controllo concerne l'insieme di dati di *input* in fase di addestramento e quello presente nel momento in cui la decisione algoritmica è stata assunta. All'uopo una proposta proveniente da esperti nel settore dei dati e dei *big data* è quella di introdurre schede tecniche rivolte agli addestratori o, più in generale, ai produttori, contenenti una serie di domande volte a fornire informazioni sul *dataset*, attinenti ai momenti salienti del ciclo di vita di un insieme di dati (*dataset lifecycle*)<sup>162</sup>, in modo da ricostruirne il contenuto.

Si potranno dare casi in cui arrestare la verifica del contributo causale del manufatto intelligente al *set* di dati potrebbe non essere soddisfacente. Per raggiungere lo standard dell'elevata credibilità razionale dell'ipotesi, dovranno essere scartati decorsi causali alternativi relativi a un eventuale ruolo causale del funzionamento del sistema che prescinda da un problema nella selezione dei dati. Questo passaggio della spiegazione causale potrà essere soddisfatto attraverso le tecniche di esplicabilità, che fanno luce sul funzionamento "interno" del sistema di IA. In alternativa, si può immaginare che l'esclusione dei fattori causali alternativi sia raggiungibile testando il sistema con i dati di cui disponeva al momento del suo utilizzo e poi con nuovi dati, privi di *bias*: se il sistema nel primo caso erra, mentre una volta inseriti dati nuovi non ripete l'errore sistematico, allora potrà escludersi il suo malfunzionamento interno e ciò confermerà l'ipotesi che l'errore che ha contribuito alla verifica dell'evento si colloca nella fase produttiva relativa alla selezione dei dati. Probabilmente bisognerà ricorrere in ogni caso alle tecniche di *explanation* laddove l'errore non sia sistematico e ci sia un problema di

---

<sup>162</sup> T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. Dumé, K. Crawford, *Datasheets for Datasets*, <https://arxiv.org/abs/1803.09010>, lo studio appare anche nel sito di Microsoft, nella sezione "Research", dove si legge: "*Datasheets for datasets is a tool for documenting the datasets used for training and evaluating machine learning models. The aim of datasheets is to increase dataset transparency and facilitate better communication between dataset creators and dataset consumers (e.g., those using datasets to train machine learning models). Datasheets encourage dataset creators to carefully reflect on the dataset creation process, enabling them to uncover possible sources of bias in their data or unintentional assumptions that they've made. For dataset consumers, the information contained within datasheets can help ensure that the dataset is the right choice for the task at hand. Datasheets can optionally be exposed to end users for increased transparency and trust. Datasheets contain questions about dataset motivation, composition, collection, pre-processing, labeling, intended uses, distribution, and maintenance. Crucially, and unlike other tools for meta-data extraction, datasheets are not automated, but are intended to capture information known only to the dataset creators and often lost or forgotten over time*".

varianza<sup>163</sup>. Trattandosi di un errore che non si presenta con regolarità, sarà più difficile da individuare e per questo potrebbe non bastare il solo esame dei dati utilizzati durante l'addestramento per affermare che la causa dell'errore risieda non nel funzionamento del sistema, ma nel tipo di addestramento ricevuto.

Queste osservazioni in tema di causalità vanno lette considerando che il giudice è tenuto ad applicare il sapere scientifico esistente al momento dell'accertamento e non quello conosciuto al momento della condotta. Contrariamente a quanto avviene in sede di accertamento dell'elemento soggettivo, dove gli sviluppi della scienza successivi alla condotta lasciano propendere per un giudizio di imprevedibilità -e quindi inevitabilità- dell'evento, "il parametro eziologico è una finestra aperta sulla conoscenza scientifica al momento del giudizio"<sup>164</sup>, per cui il giudice potrà basare il proprio prudente apprezzamento su studi e tecniche di spiegazione nuove che aiutino a far luce sulla verifica del decorso causale concreto.

Le osservazioni qui sviluppate servono per ridimensionare il problema della opacità causale e per sollecitare riflessioni che possano confluire in un apparato legislativo che imponga tecniche produttive dei sistemi intelligenti che attenuino l'oscurità del loro funzionamento. Mi sembrano, questi, propositi legislativi realizzabili, che non conosceranno importanti resistenze da parte dei produttori, poiché, vale ancora la pena ricordarlo, anche loro trarrebbero un vantaggio in termini economici dalla produzione di sistemi esplicabili e interpretabili, atteso che sistemi di tal tipo sono affidabili e quindi percepiti come più sicuri dai potenziali acquirenti.

Al contrario, l'affidabilità che il consumatore riconosce a queste nuove tecnologie non potrebbe in alcun modo essere sollecitata da una flessibilizzazione delle categorie penalistiche, in primo luogo dalla erosione della causalità in favore di un suo apprezzamento *ex ante* vista come possibile soluzione ai rischi della modernità<sup>165</sup>.

---

<sup>163</sup> Per la differenza tra errore sistematico (bias) e errore variabile (varianza) si rinvia al cap. I, par. 2.2.2.

<sup>164</sup> M. Donini, *Il garantismo della condicio sine qua non*, cit., p. 497 e 529.

<sup>165</sup> Va osservato che la tendenza a limitare l'accertamento causale al solo apprezzamento *ex ante* non era fenomeno inedito prima che si presentassero i nuovi rischi della modernità, ma era già concepita nella vicenda, tradizionale, del concorso di persone nel reato così come letta dalla teoria della prognosi o dell'aumento del rischio. All'interno di questo istituto di parte generale la flessibilizzazione del nesso causale è vista come più pericolosa da G. De Vero, *op. cit.*, p. 673, in quanto motivata da una sorta di "appetito di punibilità" che si riverserebbe in una lettura estensiva dell'art. 110 c.p., mentre la stessa rinuncia con riguardo ai nuovi danni da modernità (seppur non condivisibile) servirebbe a fronteggiare nuove forme di causalità multifattoriale.

La responsabilizzazione del consumatore dovrebbe passare anche da una visione “rigida” dello strumento penale, che non deve estendersi oltre quei limiti tracciati dalle garanzie fondamentali che arginano la tendenza espansiva del diritto del crimine, che il consumatore fatica a percepire come un’*extrema ratio*, affidandosi spesso a esso per la propria tutela. Da qui la percezione dell’opacità causale come una forma di denegata giustizia, piuttosto che come una giusta applicazione delle categorie del diritto penale d’evento.

## 5. Brevi riflessioni sulla causalità omissiva

Nei casi di danni da produzione e uso di manufatti intelligenti, la causalità che verrà in gioco sarà parimenti attiva e omissiva. Si può anche immaginare che in certi contesti queste forme di causalità si presenteranno congiuntamente e concorreranno nella causazione dell’evento, creando un contesto di concause da leggere alla luce dell’art. 41 c.p.. L’eziologia causale potrà infatti essere il risultato dell’intreccio di più condotte attive e omissive che si collocano sul piano della produzione (responsabilità del produttore) e dell’utilizzo (responsabilità dell’utente utilizzatore) e che assumono rilevanza causale se congiuntamente considerate.

L’assenza di casi già sottoposti all’attenzione della giurisprudenza non consente di analizzare la questione partendo da una base prasseologica che involge produzione e uso di tecnologie intelligenti, ma è anche vero che il dibattito dottrinale sul nesso tra omissione e evento e la casistica esemplificativa proposta negli studi sul tema, soprattutto per i casi di responsabilità medica e di malattie professionali, ci consentono già di sviluppare delle riflessioni in questa sede.

Tradizionalmente causalità attiva e “causalità omissiva”, quest’ultima riferita ai soli reati omissivi impropri<sup>166</sup>, si pongono come forme naturalistiche e concettuali distinte. Sul piano naturalistico, l’omissione è per definizione priva di una eziologia fenomenica ‘reale’, che possa nella sua materialità essere percepita come possibile causa dell’evento.

---

<sup>166</sup> Come noto, la causalità non viene in rilievo come elemento dei reati omissivi propri, il cui fatto tipico è per definizione privo di un evento e l’omissione, in quanto tale, è minacciata di pena, poiché è infranto il comando di compiere una data azione. Ma questo discorso non vale in termini assoluti: eccezionalmente anche nei reati di pura omissione si presenta la questione del rapporto di causalità quando un evento è considerato dalla legge come circostanza aggravante (si vedano la morte o le lesioni nell’omissione di soccorso, art. 593 c.p.). Sul problema della causalità dell’omissione, F. Antolisei, *Il rapporto di causalità*, cit., pp. 118 ss.

L'omissione, infatti, denota un comportamento statico, un'inerzia, che stride con la concezione di un processo causale che consti di un susseguirsi di forze cinetiche: "l'omissione è praticamente il nulla e di per sé quindi non può produrre effetti nel mondo esteriore"<sup>167</sup>.

Per questo c'è chi evidenzia la fallacia dell'espressione "causalità omissiva", atteso che una causalità sul piano naturalistico tra l'omissione e l'evento non esiste, e promuove l'impiego di formule quali "nesso tra omissione ed evento" o "equivalente normativo della causalità" o ancora "equivalente tipico della causalità"<sup>168</sup>.

Ma attenzione, queste affermazioni non valgono ad annullare il decorso reale che causa l'evento: in parallelo alla condotta omissiva vi è un'eziologia materiale e naturalistica che si sviluppa sul piano reale e che conduce alla verifica dell'evento<sup>169</sup>. È questo il caso dell'insorgenza della malattia che porta alla morte, ovvero dell'esposizione ad amianto che determina lo sviluppo della patologia tumorale<sup>170</sup> o del funzionamento del sistema di IA da cui scaturisca un'offesa.

Il paradigma omissivo improprio può allora essere immaginato tracciando due rette parallele: da un lato quella del decorso reale, che è la causa naturale dell'evento, suscettibile di una verifica in concreto; dall'altro l'equivalente normativo della causalità, non riscontrabile empiricamente poiché avente natura ipotetica. Le due rette non si incontrano se non al momento dell'accertamento della rilevanza della condotta doverosa omessa: l'accertamento della responsabilità omissiva postula l'interruzione del tracciato parallelo seguito dalle due rette, per verificare gli sviluppi che sarebbero sorti dalla loro intersezione, al fine di comprendere l'efficacia impeditiva del comportamento alternativo lecito.

---

<sup>167</sup> *Ibidem*, p. 120.

<sup>168</sup> P. Veneziani, *Il nesso tra omissione ed evento*, cit., p. 1974; ma già G. Grasso, *Il reato omissivo improprio*, Milano, Giuffrè, 1983, p. 417. Nota però l'A. che l'espressione "causalità omissiva" sembrerebbe essere promossa dallo stesso art. 40, c.1, laddove unifica l'azione e l'omissione, oltre al fatto che in entrambi i casi si ricorre all'impiego di leggi scientifiche per l'accertamento del nesso di condizionamento. Infine, alcuni reati commissivi presentano un legame a carattere ipotetico tra comportamento e risultato, portando a tal proposito l'esempio della "catena causale di salvataggio", in cui si debba accertare il rapporto causale tra la condotta di A, che impedisce a B di salvare C, e la morte di C: in questo caso non si pone in discussione il termine "causalità", sebbene si tratti di una condotta commissiva che necessita un accertamento prognostico per verificare come si sarebbero svolti i fatti senza l'intervento di A.

<sup>169</sup> P. Veneziani, *op. cit.*, p. 1975; G. Marinucci, *Causalità reale e causalità ipotetica nell'omissione impropria*, in *Riv. it. dir. proc. pen.*, 2009, pp. 523 ss.

<sup>170</sup> S. Zirulia, *op. cit.*, p. 62.

Va certamente dato atto dei tentativi di rivestire di realtà l'equivalente tipico della causalità, attraverso la valorizzazione dell'*aliud factum* o *actum* (l'azione che il soggetto compie mentre omette l'atto comandato)<sup>171</sup>, ovvero dell'elaborazione dell'inerzia come processo statico o condizione statica 'reale'<sup>172</sup>, ma si tratta di tentativi superati o, addirittura, abbandonati dagli stessi sostenitori di questo pensiero<sup>173</sup>; a prevalere è oggi la tesi per cui l'omissione non può *causare nulla*<sup>174</sup>.

L'inconsistenza naturalistica dell'omissione la distingue allora dall'azione e questa differenza si riflette sul piano giuridico-concettuale, che è il terreno in cui il nesso tra omissione e evento prende forma, acquisendo una sostanza normativa, contrapposta alla materialità della causalità commissiva. Nel nostro ordinamento è lo stesso codice, all'art. 40 cpv., a fornire gli strumenti per definire la causalità omissiva come *equivalente* normativo di quella attiva, posto che "non impedire un evento che si ha l'obbligo giuridico di impedire *equivale* a cagionarlo". La causalità omissiva, inesistente sul piano reale, si afferma su quello normativo, cosicché è il concetto a creare il fenomeno e non l'inverso: una condotta omissiva è causa dell'evento se, immaginando come realizzato il comportamento omesso, l'evento non si sarebbe verificato. Da qui l'applicazione della teoria condizionalistica "modificata" da un giudizio *contra factum* di "aggiunta" piuttosto che di "eliminazione" mentale<sup>175</sup>. Ne segue un enunciato causale del tipo "se la condotta doverosa fosse stata tenuta, l'evento non si sarebbe verificato". Si tratta di un enunciato ipotetico (in grammatica, un periodo ipotetico dell'irrealtà), definito in logica come "condizionale controfattuale", la cui protasi enuncia un'ipotesi contraria a quanto è realmente accaduto, mentre l'apodosi enuncia la conseguenza che sarebbe derivata da quell'ipotesi<sup>176</sup>. Il condizionale controfattuale non ha ad oggetto né la spiegazione di un

---

<sup>171</sup> Questo, come ricorda Antolisei, *cit.*, p. 122, è il primo momento del pensiero del Luden in *Abhandlungen aus dem gemeinen deutschen Strafrecht*, vol. I, 1836 pp. 467 ss. e vol. II, 1840, pp. 219 ss., che è il primo autore a sollevare il problema della causalità dell'omissione.

<sup>172</sup> F. Stella, *La nozione penalmente rilevante di causa: la condizione necessaria*, in *Riv. it. dir. proc. pen.*, 1988, pp. 1249 ss.

<sup>173</sup> È il Luden ad abbandonare la teoria dell'*aliud agere* in una seconda fase del suo pensiero, come ricorda F. Antolisei, *op. cit.*, p. 122 nota (4), il quale a p. 124 osserva anche che la tesi è smentita dal fatto che il luogo in cui si verifica l'omissione non è quello in cui si espleta la diversa condotta attiva, bensì quello in cui si sarebbe dovuto compiere l'azione prescritta. Questo dimostra che la condotta diversa tenuta in luogo dell'agire doveroso non ha alcuna rilevanza per il diritto, al punto tale che il giudice non deve neppure accertarla. Sulla irrilevanza dell'*aliud agere* in ragione della natura normativa dell'omissione, per cui assume importanza solo ciò che si sarebbe dovuto fare e non si è fatto, anche M. Romano, *Pre-art. 39*, in *Commentario*, *cit.*, p. 313.

<sup>174</sup> G. Fiandaca, *voce Causalità*, in *Dig. Disc. Pen.*, II, Torino, UTET, 1988, pp. 126-127; G. Grasso, *op. cit.* pp. 385 ss.

<sup>175</sup> Sul punto, G. Grasso, *op. cit.*, p. 387 ss.

<sup>176</sup> *Voce Condizionale controfattuale* in *Dizionario di filosofia*, 2009, [www.treccani.it](http://www.treccani.it).

evento nel passato, né la previsione di un accadimento futuro, e si caratterizza in quanto l'antecedente e il conseguente sono falsi per definizione, posto che "l'azione doverosa non è stata compiuta" (smentita dell'antecedente) e l'evento si è verificato (smentita del conseguente)<sup>177</sup>. Il valore di verità del condizionale controfattuale, allora, dipende da una connessione causale che intercorre tra il fatto ipotizzato nella protasi e la conseguenza enunciata nell'apodosi, che è ricavata da una generalizzazione causale enunciata in una legge scientifica: per questo si è correttamente detto che anche il condizionale controfattuale postula come necessario il ricorso al modello della sussunzione sotto leggi<sup>178</sup>, in modo analogo a quanto avviene nel modello di spiegazione delle condotte attive. Ma a ben vedere, la struttura controfattuale dell'enunciato condizionalista non è una caratteristica esclusiva dell'omissione: anche la spiegazione causale nei reati commissivi ricorre alla struttura logica del condizionale controfattuale per effetto della formula della *condicio*<sup>179</sup>. Nondimeno, diverso è il ruolo che svolge il nesso tra la condotta e l'evento all'interno della fattispecie nei reati commissivi e omissivi. Nei primi, la causalità serve a selezionare il comportamento antigiuridico; nell'omissione, il comportamento di cui si deve valutare la rilevanza penale non è quello legato all'evento lesivo dal nesso eziologico (che manca in *rerum natura*), bensì quello individuabile attraverso la situazione di garanzia che, tra i vari comportamenti impeditivi dell'evento, ne seleziona uno. Nei reati omissivi impropri la "causalità" perde allora quel ruolo di "filtro" che essa normalmente svolge nelle corrispondenti fattispecie commissive<sup>180</sup>: solo se esiste un obbligo giuridico di impedire l'evento, il non impedimento dell'evento equivale alla sua causazione<sup>181</sup>.

Un'altra differenza dell'enunciato controfattuale dell'omissione risiede nella sua natura doppiamente ipotetica che emerge dal confronto tra la protasi dell'enunciato controfattuale omissivo e la protasi di quello attivo<sup>182</sup>. Infatti, come nota Paliero, mentre

---

<sup>177</sup> G. Grasso, *op. cit.*, p. 387.

<sup>178</sup> G. Grasso, *op. cit.*, p. 392.

<sup>179</sup> F. Stella, *La condizione penalmente rilevante di causa; la condizione necessaria*, in *Riv. it. dir. proc. pen.*, 1988, p. 1254, il quale legge in modo critico la conclusione di Grasso sul punto; C.E. Paliero, *La causalità dell'omissione: formule concettuali e paradigmi prasseologici*, in *Riv. it. med. Leg.*, XIV, 1992, p. 839-840, il quale ricostruisce il pensiero dei due autori, per sposare (ma solo) su questo punto le conclusioni di Stella.

<sup>180</sup> G. Grasso, *op. cit.*, pp. 416-417.

<sup>181</sup> È il pensiero di F. von Liszt, il quale sposa una concezione radicale, come ricorda Antolisei, *op. cit.*, p. 138, secondo la quale la punibilità dell'omissione è indipendente dalla sua causalità: all'azione causale si contrappone l'omissione non causale. Listz, *Lehrbuch des deutschen Strafrechts*, De Gruyter, 1894 pp. 108 ss..

<sup>182</sup> C.E. Paliero, *op. cit.*, p. 842.

il *criterio* del giudizio è imperniato su un sillogismo controfattuale, sia nell'azione che nell'omissione, diversa è la *base* di tale giudizio, cioè l'antecedente oggetto della protasi controfattuale<sup>183</sup>.

E in effetti, notiamo che i due condizionali controfattuali coincidono quanto ad apodosi (che in entrambi i casi è "l'evento non si sarebbe verificato"), ma divergono nella protasi che non è doppiamente ipotetica nel condizionale controfattuale tra condotta attiva ed evento, in quanto si ipotizza come non realizzata la condotta tenuta dall'agente, sostituendola con l'inerzia.

Nel controfattuale dell'omissione, l'apodosi consta di due componenti ipotetiche: la prima è la condotta omessa dal garante, che il giudice ipotizza come realizzata sulla base del giudizio controfattuale; la seconda attiene all'efficacia del comportamento non tenuto di impedire l'evento. Questo secondo momento ipotetico rende l'omissione un "concetto di relazione", che assume significato condizionante solo se correlato a un modello di condotta attiva specifico, appunto il comportamento alternativo lecito<sup>184</sup>.

### **5.1 Alcuni aspetti problematici. La difficile distinzione tra azione e omissione**

La distinzione naturalistica e concettuale tra causalità attiva e omissiva non vale certo a superare quanto abbiamo detto finora, perché l'impiego delle leggi scientifiche anche per la causalità omissiva comporta che essa presenti tutti i problemi di quella attiva<sup>185</sup>; tuttavia, in aggiunta porta con sé problemi ulteriori.

Un primo aspetto problematico concerne l'obbligo di attivarsi, quindi la situazione di garanzia quale elemento del fatto tipico che, nei reati omissivi impropri, si interseca e in parte si sovrappone col dovere di diligenza, ma che non arriva a identificarsi totalmente con esso<sup>186</sup>. La posizione di garanzia rientra nella fattispecie oggettiva dell'omissione e la sua esistenza prescinde dal tema qui in esame, cioè la causalità. Nondimeno, questo elemento della fattispecie giunge a condizionare l'elemento causale nel momento in cui l'accertamento di quest'ultimo venga nella prassi fatto coincidere con la sussistenza della posizione di garanzia, che è la sola a essere accertata. La dottrina descrive questo

---

<sup>183</sup> *Ibidem*, p. 841-842.

<sup>184</sup> *Ibidem*, p. 843.

<sup>185</sup> Sull'identità strutturale della spiegazione ex post nella causalità attiva e omissiva e sulla necessità di un accertamento con l'impiego di leggi di copertura, M. Romano, *Commentario, cit. sub art. 40/36 ss.*, pp. 375 ss.

<sup>186</sup> G. Grasso, *op. cit.*, p. 372 ss.

fenomeno come “fuga in avanti” nell’accertamento della fattispecie, per indicare l’appiattimento della (preliminare) prova del nesso causale nella (successiva) dimostrazione della sussistenza di una posizione di garanzia<sup>187</sup>. L’atteggiamento è sicuramente da criticare: l’obbligo giuridico di neutralizzare l’evento non comporta *ex se* la rilevanza causale della condotta.

Un secondo momento problematico è quello dell’efficacia impeditiva della condotta doverosa omessa. Questo effetto impeditivo va verificato al fine di determinare la rilevanza penale dell’inerzia: solo la causazione *evitabile* di eventi determina il disvalore penale di una condotta che diversamente non sarebbe meritevole di pena<sup>188</sup>.

L’accertamento dell’efficacia impeditiva del comportamento alternativo lecito pone delle criticità nella prassi, dove è difficile distinguere tra azione e omissione quando entra in gioco il reato colposo. È qui che i segmenti dell’omissione causale coincidono col segmento omissivo della colpa, cioè con l’inosservanza del comportamento diligente imposto dalla regola cautelare.

Il punto di partenza di questo ragionamento è l’efficacia impeditiva della condotta omessa. Il problema che presenta questa seconda componente ipotetica del giudizio ex art. 40 cpv. c.p. relativo all’omissione è legato alla natura astratta del giudizio: non essendo concretamente stata posta in essere la condotta doverosa, la sua efficacia impeditiva è valutata solo in via ipotetica. È proprio su questo punto che si instaura una corrispondenza tra il nesso tra omissione ed evento e la componente omissiva della colpa, che rende poco chiaro il confine tra causalità attiva e omissiva<sup>189</sup>.

Nel reato colposo l’accertamento del fatto tipico non si arresta al controllo di congruità tra l’evento concretamente verificatosi e quello che la regola di diligenza mirava a prevenire, ma ci si spinge anche ad accertare la capacità preventiva di tale regola e, quindi, l’attitudine della condotta cautelare prescritta a evitare l’evento: si accerta *ex post* il fallimento della regola di diligenza, quest’ultima prospettata *ex ante*, nel contrastare il rischio identificato *ex ante* e realizzatosi nell’accadimento concreto<sup>190</sup>. In questo modo due sono i momenti di “particolarizzazione della colpa sull’evento”, cioè

---

<sup>187</sup> C.E. Paliero, *op. cit.*, pp. 830-831.

<sup>188</sup> L. Eusebi, *Appunti sul confine tra dolo e colpa, cit.*, p. 1063.

<sup>189</sup> M. Donini, *La causalità omissiva, cit.*, pp. 41 ss.; R. Blaiotta, *La ricostruzione del nesso causale nelle esposizioni professionali*, in *Cass. pen.*, 2, 2006, pp. 800 ss; K. Summerer, *Causalità ed evitabilità. Formula della condicio sine qua non e rilevanza dei decorsi causali ipotetici nel diritto penale*, Pisa, Edizioni ETS, 2013, pp. 138 ss.

<sup>190</sup> G. Forti, *Colpa ed evento nel diritto penale*, Milano, Giuffrè, 1990, pp. 663-665.

quello della realizzazione del rischio e quello dell'evitabilità dell'evento per effetto del comportamento alternativo lecito<sup>191</sup>.

Orbene, così come l'accertamento causale nei reati omissivi, anche il momento omissivo della colpa (o la causalità del comportamento alternativo lecito) postula un ragionamento ipotetico di aggiunta, poiché l'interprete è chiamato a immaginare il decorso causale che si sarebbe verificato se fosse stata tenuta la condotta conforme omessa. Ciò implica che il reato omissivo e quello attivo colposo non rappresentano forme separate di costruzione della fattispecie penale poiché la realtà che essi descrivono è la medesima: l'inosservanza di un obbligo da parte di un certo soggetto. La differenza è la fonte dell'obbligo, che è la sussistenza di una particolare posizione di garanzia nell'omissione e lo svolgimento di un certo tipo di attività consentita nel reato colposo<sup>192</sup>.

Bisogna però capire se questi due segmenti omissivi occupano lo stesso posto nella struttura dell'illecito omissivo e di quello colposo, ovvero se si collocano su piani diversi.

Alla luce di quanto sinora detto in tema di omissione, non v'è dubbio che il comportamento doveroso omesso svolge un ruolo all'interno della categoria della causalità dei reati omissivi: anzi, esso esaurisce il giudizio causale dell'omissione impropria, che inizia con l'individuazione della causa naturale dell'evento e termina col procedimento dell'eliminazione mentale per controllare che l'omissione sia condizione dell'evento (questi i due elementi della causalità omissiva definiti come "il prima" e "il poi"<sup>193</sup>). Come detto, questo accertamento utilizza il metodo della sussunzione del caso sotto le leggi scientifiche, che è una sussunzione doppia, una per la causa naturale, l'altra per l'efficacia impeditiva del comportamento alternativo lecito<sup>194</sup>.

Appurato il ruolo per definizione causale del comportamento alternativo lecito nell'omissione impropria, più discussa è invece la collocazione dello stesso nel reato commissivo colposo, nello studio del quale si registrano opinioni tese a collocarlo nella categoria della causalità e in quello della colpa.

Secondo la prima visione dottrinale, vi è piena coincidenza tra la nozione di causalità cui fanno riferimento gli artt. 40 e 43 c.p., con riguardo l'uno al rapporto dell'evento con la condotta, l'altro al rapporto tra l'evento e la violazione di una regola cautelare<sup>195</sup>. Ne

---

<sup>191</sup> *Ibidem*, p. 668.

<sup>192</sup> L. Eusebi, *op. cit.*, pp. 1064-65.

<sup>193</sup> G. Marinucci, *op. cit.*, p. 523.

<sup>194</sup> F. Mantovani, *Diritto penale, cit.*, p. 154.

<sup>195</sup> L. Eusebi, *op. cit.*, p. 1061, p. 1066.

deriva che in ogni caso l'impedimento dell'evento da parte del comportamento alternativo lecito deve essere accertato con una probabilità vicina alla certezza.

La seconda tesi, che ritaglia uno spazio nella (dimensione oggettiva della) colpa, evita proprio questa conseguenza e sottrae l'accertamento dell'efficienza causale ipotetica della condotta diligente omessa alla rigidità dell'accertamento causale, così sostituendo alla certezza processuale l'impiego di valutazioni probabilistiche e della più morbida costruzione della mancata diminuzione del rischio<sup>196</sup>.

L'opinione da ultimo riferita ha delle ricadute sulla responsabilità omissiva, che si conferma come elemento dell'illecito da accertare secondo i parametri dell'alta credibilità razionale o della probabilità confinante con la certezza processuale. Se infatti si argomenta l'impiego della teoria dell'aumento del rischio per l'accertamento del comportamento alternativo omesso alla luce della sua rilevanza in sede di accertamento della componente oggettiva della colpa anziché della causalità, non si può poi tollerare nessun affievolimento in punto di causalità omissiva, dove il problema dell'evitabilità dell'evento per effetto del comportamento doveroso non tenuto diventa un problema causale e pretende perciò la ragionevole certezza propria della causalità<sup>197</sup>. Questa rigidità nella costruzione della causalità omissiva è del resto propugnata dalla stessa dottrina, che critica l'atteggiamento giurisprudenziale che mirava ad accentuare il momento omissivo della condotta (attiva) colposa intesa come omessa diligenza, proprio allo scopo di agevolare la prova della causalità, in ragione del fatto che quella omissiva potesse essere meno certa e più probabilistica<sup>198</sup>. Infatti, valorizzando il momento omissivo della condotta colposa, l'accertamento della capacità del comportamento doveroso omesso di impedire l'evento, cioè il rischio tipico che la norma cautelare mirava ad evitare, potrebbe avvenire sotto la lente delle teorie dell'imputazione obiettiva dell'evento, in particolare dell'aumento del rischio. Procrastinando siffatto accertamento al momento omissivo della colpa si aggirerebbe l'ostacolo della certezza processuale circa l'efficacia impeditiva del comportamento alternativo lecito, poiché la cosiddetta "causalità della colpa" non è elemento del nesso di condizionamento di cui all'art. 40 c.p., bensì un aspetto della (dimensione oggettiva della) colpa come definita dall'art. 43 c.p., cioè un comportamento

---

<sup>196</sup> M. Donini, *op. cit.*, pp. 75 ss.

<sup>197</sup> R. Blaiotta, *op. cit.*, p. 801.

<sup>198</sup> M. Donini, *op. cit.*, pp. 75 ss.; R. Blaiotta, *op. cit.*, p. 801.

che determina il prodursi dell'evento *a causa* di negligenza o imprudenza o imperizia, ovvero per inosservanza di leggi, regolamenti, ordini o discipline.

L'eccessiva flessibilizzazione dei criteri impiegati per l'accertamento della causalità omissiva è stata alla base del contrasto che ha poi portato alla nota pronuncia delle Sezioni Unite nel 2002. Oggi anche l'eziologia omissiva va accertata secondo i canoni delle Sez. Un. Franzese, pronunciatesi proprio su un caso di responsabilità (medica) per omissione<sup>199</sup>. Valgono, allora, le conclusioni raggiunte in punto di leggi statistiche utilizzabili e accertamento causale secondo i criteri della probabilità logica o elevata credibilità razionale<sup>200</sup>.

È, questo, il risultato di compromesso tra gli orientamenti contrastanti registratisi nella nostra giurisprudenza che oscillava tra applicazioni surrettizie della teoria dell'aumento o mancata diminuzione del rischio<sup>201</sup>, da un lato, e tentativi di recupero della certezza del condizionamento causale, invocando anche per l'omissione l'impiego di leggi illustrative di regolarità causali vicine a alla certezza, cioè prossime a 100<sup>202</sup>.

La pronuncia delle Sezioni Unite supera questo contrasto indicando una "terza via": come il secondo degli orientamenti sopra citati, afferma un criterio di accertamento omogeneo tra causalità attiva e omissiva, così sbarrando la strada all'ingresso di letture dei reati di evento come reati di pericolo attraverso la debolezza del nesso di derivazione e favorendo una ricostruzione unitaria della causalità; essa apre verso l'utilizzo di leggi statistiche con frequenze anche medio-basse, come ricordato poco sopra, valorizzando la certezza processuale (raggiunta attraverso la prova per esclusione) piuttosto che quella scientifica (che postula l'impiego di leggi scientifiche universali).

In questo modo la giurisprudenza conferma quanto la dottrina aveva già osservato, cioè la necessità di un accertamento causale con una probabilità vicina alla certezza processuale anche in presenza di un'inerzia<sup>203</sup>. Potremmo allora dedurne un superamento di quell'approccio teso a valorizzare l'inosservanza della regola cautelare come omissione, posto che la certezza (processuale) è un requisito dell'accertamento del nesso che lega l'evento tanto all'azione, quanto all'omissione. Anzi, si potrebbero addirittura

---

<sup>199</sup> Per lo studio della pronuncia nella specifica costellazione del reato omissivo improprio, si rinvia a P. Veneziani, *Il nesso tra omissione ed evento nel settore medico*, cit., pp. 1969 ss.

<sup>200</sup> M. Romano, *Commentario*, cit., sub art. 40/43, p. 378.

<sup>201</sup> *Ex multis*, Cass. sez. IV, 12 luglio 1991, Silvestri, in *Foro it.*, 1992, p. 363.

<sup>202</sup> Ad es. in Cass., sez. IV, 28 settembre 2000, Baltocchi, in *Cass. pen.*, 2002, pp. 159 ss.

<sup>203</sup> M. Donini, *op. cit.*, p. 75.

porre le basi per un fenomeno opposto, che valorizzi la causalità colposa come momento dell'agire colposo ex art. 43 c.p. -e non come omissione causale ex art. 40 cpv.- che consente un accertamento “a maglie larghe” sulla probabilità di evitare l'evento inteso come mancata diminuzione delle chances di salvezza del bene protetto. In altri termini, dopo che nel 2002 le Sezioni unite hanno chiarito che anche per la causalità omissiva vale il criterio della probabilità logica e non v'è quindi spazio per la teoria della mancata diminuzione del rischio, non v'è motivo per leggere l'inosservanza del dovere di diligenza come condotta omissiva piuttosto che qualificarla correttamente come commissiva. Il fatto colposo non va allora qualificato come violazione di un comando che impone di osservare la diligenza (avendo così sempre una struttura omissiva), ma piuttosto come “un'azione evitabile lesiva del bene giuridico vietata dall'ordinamento”<sup>204</sup>.

Nondimeno, siffatta conclusione sarebbe affrettata, poiché non considererebbe le difficoltà che tuttora persistono per la prova della causalità nell'omissione, che si presentano quindi in caso di qualificazione in chiave omissiva della condotta. Il quadro che emerge sembrerebbe non lasciare spazio all'incertezza in ordine all'efficacia impeditiva del comportamento alternativo omesso, che è un aspetto che la giurisprudenza è chiamata ad affrontare sia in sede di causalità, che in sede di accertamento della colpa, se non vuole cadere nella tentazione di accontentarsi di un grado di accertamento inferiore rispetto a quello richiesto per le condotte in cui un'omissione non si verifica, così rendendosi artefice del processo di erosione della causalità sul terreno del diritto penale d'evento.

Al di là della ritrosia riscontrata nella giurisprudenza nel valorizzare la componente attiva di condotte colpose che vengono invece qualificate come omissive<sup>205</sup>, che quindi marginalizza le osservazioni che seguono, non vanno sottovalutati i rischi di una qualificazione rimessa alla libera interpretazione giurisprudenziale e più in particolare della pubblica accusa, la quale potrebbe prediligere quella tra le due categorie (reato colposo omissivo improprio o reato colposo commissivo) che secondo l'orientamento seguito dall'opinione dominante conduce alla condanna tramite un sentiero meno tortuoso. In altri termini, il dilemma tra la qualificazione della condotta come attiva o omissiva non dovrebbe essere sciolto facendosi guidare dalle difficoltà probatorie che meno ostacolano una risposta punitiva reputata come necessaria per via dei nuovi rischi che caratterizzano

---

<sup>204</sup> G. Forti, *op. cit.*, p. 181.

<sup>205</sup> R. Blaiotta, *op. cit.*, p. 276.

la società moderna. Per questo un importante ruolo è quello della dottrina: da un proficuo dialogo tra tecnici del diritto e giuristi potrà certamente scaturire un atteggiamento più attento nella qualificazione di una condotta come attiva ovvero omissiva. Si tratta, del resto, di un esperimento già riuscito in alcuni contesti tematici, in particolare nel settore della responsabilità dell'imprenditore per gli infortuni o le patologie insorte a causa dello svolgimento dell'attività lavorativa, come nota la dottrina che si è occupata dell'argomento<sup>206</sup>.

Proprio al fine di favorire questo dialogo tra dottrina e giurisprudenza nel settore della responsabilità medica, un'opinione valorizza la distinzione tra il rischio creato o incrementato dal reo, che sarebbe alla base di una qualificazione in senso attivo della condotta, da un canto, e il rischio non evitato, dall'altro, dove è il decorso naturalistico a produrre l'evento, che poteva essere impedito dall'autore di una condotta qualificata come omissiva<sup>207</sup>. Più in generale, si osserva che bisogna avere riguardo alla "paternità del fatto"<sup>208</sup>, per cui vi sarà azione quando è la condotta positiva a innescare il processo causale che porta all'evento; mentre in presenza di una condotta omissiva l'evento è cagionato da forze esterne, naturali o umane (o di agenti artificiali, aggiungiamo noi), e la neutralizzazione dell'evento dà consistenza all'obbligo che grava sul garante<sup>209</sup>.

Ne deriva che in un contesto produttivo, la responsabilità dell'imprenditore per i danni cagionati dal prodotto difettoso è conseguenza di una condotta attiva, consistente nella collocazione del prodotto sul mercato. Qui, in effetti, viene violata una regola di diligenza che deve governare l'attività produttiva, cioè quella che impone la produzione di prodotti sicuri.

Mi sembra invece qualificabile come omissiva la condotta dell'utilizzatore che ometta un intervento impeditivo di un danno causato da un errore del sistema di IA: la situazione di rischio è stata *creata* dalla condotta attiva del produttore ed è stata *incrementata* dall'utilizzatore a causa di una mancata manutenzione del prodotto o di un'azione tempestiva che potrebbe consistere in una ripresa in mano del controllo di una macchina (ad esempio un'auto a guida autonoma) comandata da un algoritmo. A una simile ricostruzione in chiave omissiva si potrebbe obiettare che comunque è

---

<sup>206</sup> S. Zirulia, *op. cit.*, p. 63 sottolinea la maggiore attenzione della più recente giurisprudenza nel qualificare come attiva la condotta del datore di lavoro per la mancata adozione di cautele nei luoghi di lavoro nelle patologie amianto-correlate.

<sup>207</sup> P. Veneziani, *op. cit.*, p. 1972.

<sup>208</sup> R. Blaiotta, *op. cit.*, p. 276; Donini, *op. cit.*, p. 44.

<sup>209</sup> K. Summerer, *op. cit.*, p. 141.

l'utilizzatore ad attivare il sistema, "a premere il pulsante di accensione". Ciò è indiscutibile, poiché alla base vi è una condotta attiva, ma se poi a essa segue un processo di funzionamento totalmente indipendente del sistema, che procede autonomamente senza alcun intervento umano e senza che questo intervento sia richiesto al di là dei casi di emergenza, ecco che un'eventuale offesa penalmente rilevante risulta cagionata da una forza estranea al reo, che è la "forza artificiale", i cui sviluppi non sono stati neutralizzati da chi aveva l'obbligo di farlo, cioè l'utilizzatore. A ritenere diversamente, cioè a insistere sulla valorizzazione del momento in cui l'utilizzatore ha attivato il sistema, si giungerebbe a negare spazio all'omissione perché, a rigore, risalendo la catena causale si nota che ancora più a monte della condotta di uso del bene vi è pur sempre una condotta attiva, cioè l'acquisto dello stesso. Così opinando ci si inserisce in un ragionamento a ritroso simile a quello che sta alla base della critica denominata "del regresso all'infinito" mossa alla teoria condizionalistica, che giunge a ritenere causale anche la procreazione dell'omicida da parte della madre. Ciò che sfugge a questa obiezione è la descrizione dell'evento da considerarsi in concreto, *hic et nunc*. Per ritornare alle situazioni di nostro interesse, se l'evento è stato causato da un errore nel funzionamento del sistema che richiedeva una ripresa del controllo da parte dell'utilizzatore, qui l'evento è il risultato di un'inerzia dell'utilizzatore, che si aggiunge alla condotta commissiva colposa del produttore, il quale ha collocato nel mercato un bene difettoso.

Vi potranno poi essere situazioni più difficili da qualificare. Si pensi al caso del mancato ritiro del prodotto dal mercato nonostante eventuali difetti denunciati, che potrebbe essere considerato come una prosecuzione dell'*azione* di produzione e di immissione del prodotto sul mercato ovvero come una nuova dinamica comportamentale, originata dalla condotta attiva di vendita del prodotto, ma sganciata da essa e quindi vista come un comportamento omissivo<sup>210</sup>.

## 5.2 (segue) La prova per esclusione

Occupiamoci adesso brevemente della persistente problematicità dell'accertamento del nesso tra omissione ed evento anche all'indomani della pronuncia del 2002.

---

<sup>210</sup> È in questo secondo senso che i giudici tedeschi hanno giudicato il mancato ritiro dal mercato del prodotto nel caso *Lederspray*, invocando la posizione di garanzia *ex Ingerenz*, in cui la responsabilità per omissione deriva dall'azione precedente pericolosa, v. Piergallini, *op. cit.*, p. 54.

L'omogeneità del criterio di accertamento professata dalla sentenza Franzese non elimina infatti le evidenti peculiarità dell'omissione, dove "è il fittizio a dominare la scena"<sup>211</sup>, in quanto è da accertare l'efficacia impeditiva di un'azione che non c'è stata. Nella ricostruzione causale si prende in considerazione la condotta omessa, cioè il comportamento alternativo lecito non tenuto dall'agente, ed è a questa che si riferisce la legge di copertura. La certezza sull'efficacia causale, ove raggiunta, sarebbe comunque sempre una certezza "pensata", poiché si ragiona in termini ipotetici.

Così è la natura stessa dell'omissione causale a creare delle crepe nella lettura dei principi delle Sezioni Unite del 2002 e a rendere problematico il raggiungimento della probabilità logica, attraverso l'esclusione dei fattori causali alternativi, criterio ritenuto non applicabile, cosicché c'è chi ritiene che i principi della pronuncia Franzese siano strutturalmente inadeguati<sup>212</sup>.

In particolare, l'impossibilità di ricorrere in concreto al positivo riscontro del fatto che è proprio la condotta dell'imputato ad aver cagionato l'evento con l'eliminazione dei decorsi alternativi sarebbe una conseguenza della natura doppiamente ipotetica del nesso tra omissione ed evento. Si sostiene che la prova per esclusione possa funzionare solo se il fattore di cui va vagliata la rilevanza (cioè il contributo dell'imputato) si è realmente verificato e ciò non vale per la condotta omessa; inoltre, si osserva, nel reato omissivo improprio esiste per definizione un fattore causale alternativo alla condotta incriminata in grado di cagionare l'evento, che è del resto la causa naturalistica dell'evento (ad esempio, in caso di omissione del medico, il fattore alternativo che porta alla morte del paziente è la malattia). Il risultato è che il ragionamento delle Sezioni unite fallisce proprio nel settore in cui il caso concreto si inseriva, cioè quello dei reati omissivi impropri<sup>213</sup>.

Le difficoltà probatorie delle fattispecie omissive improprie sono emerse anche nella giurisprudenza post Franzese, dove talvolta si è giunti a un accertamento più flessibile della causalità, celato dalle formule linguistiche dell'elevata credibilità razionale o probabilità logica cui si presta un'adesione solo formale<sup>214</sup>.

---

<sup>211</sup> M. Romano, *Nesso causale*, cit., p. 912.

<sup>212</sup> D. Pulitanò, *Diritto penale*, cit., p. 192; P. Veneziani, *Il nesso tra omissione ed evento nel settore medico: struttura sostanziale ed accertamento processuale*, in *Studi in onore di Giorgio Marinucci*, E. Dolcini, C.E. Paliero (a cura di), p. 1970; L. Maserà, *Il modello causale delle Sezioni unite e la causalità omissiva*, in *Dir. pen. proc.*, 2006, 4, pp. 493 ss.

<sup>213</sup> L. Maserà, *op. cit.*

<sup>214</sup> È questa la critica mossa alla pronuncia della Cass. pen., Sez. IV, 2076 del 15.10.2002, *Loi sulla morte di alcuni paracadutisti durante l'utilizzo del paracadute*, da F. D'Alessandro, *Codice penale commentato*, cit., sub art. 40, p. 429.

La causalità omissiva resta allora un terreno scivoloso anche dopo la pronuncia delle Sezioni Unite, la quale continua però a essere il punto di riferimento nell'accertamento del nesso di condizionamento nel nostro ordinamento. Questa longevità della sentenza Francese è stata ricondotta alla sua duttilità che, se è vero essere la causa di un utilizzo solo retorico dei principi in essa enunciati, rende la causalità fortemente reattiva al contesto. Ne deriva che l'inadeguatezza dell'eliminazione dei decorsi alternativi in un contesto omissivo non va generalizzata e anzi, in un ambito molto problematico, quale è quello della responsabilità medica, si potrebbero porre casi in cui le difficoltà (*rectius* impossibilità) riscontrate per le omissioni caratterizzino invece condotte attive, e ciò dimostrerebbe che l'incertezza non è tanto legata al fatto commissivo o omissivo, ma alle sue caratteristiche, cioè al tipo di evento-malattia<sup>215</sup>.

Una simile conclusione può essere utile per la futura analisi dei casi che la diffusione di sistemi intelligenti porterà all'attenzione delle corti, laddove l'offesa al bene tutelato richieda una risposta penale. L'approccio dovrà essere casistico: le difficoltà di tipo logico che si oppongono alla prova per esclusione nella causalità omissiva non devono per ciò solo determinare una rinuncia al raggiungimento della certezza pratica della rilevanza causale della condotta. Così vi potranno essere dei casi in cui anche in presenza di una condotta omissiva si potrà affermare la responsabilità dell'imputato con certezza<sup>216</sup>: si pensi al caso del freno di emergenza non attivato dal conducente di un'auto a guida autonoma che avrebbe evitato l'impatto col pedone che attraversava incautamente una strada lontano dalle strisce pedonali. Più complessi saranno i casi di responsabilità all'interno della filiera produttiva, per le quali si porrà il problema di escludere fattori causali alternativi, al di là della qualificazione di tali condotte come attive o omissive: ad esempio, la certezza in ordine al fatto che l'errore del sistema sia dovuto al malfunzionamento di alcune componenti del manufatto, o a errori nell'assemblaggio delle parti del prodotto si potrà avere escludendo errori nella selezione del set di dati di *input*, con le difficoltà che tale verifica comporta a causa del segreto industriale e del complesso processo di interpretazione della decisione algoritmica.

---

<sup>215</sup> O. Di Giovine, *Lo statuto epistemologico (?) della causalità penale fra cause necessarie e condizioni sufficienti*, in *Il diritto penale di fronte alle sfide della "società del rischio"*, cit., pp. 213-214.

<sup>216</sup> Questa è anche l'osservazione di G. Fiandaca e E. Musco, *Diritto penale*, cit., p. 641, che notano che anche in un contesto omissivo si potrà raggiungere la certezza (e non la mera probabilità confinante con la certezza), come accadrebbe nell'accertare la responsabilità di una baby sitter lasciata che un bambino giochi nei pressi di uno stagno nelle cui acque da lì a breve cadrà perdendo la vita, commentando che la dovuta sorveglianza della donna sarebbe bastata a scongiurare con certezza l'evento letale.

## 6. Nuove prospettive

Lo studio delle nuove tecnologie si rivela un'occasione per continuare a indagare gli sviluppi del diritto penale e il suo rapporto con il processo. Dalle riflessioni condotte in queste pagine emerge la centralità dell'accertamento processuale e il limite delle categorie dogmatiche, che risultano sterili se non seguite da un'applicazione seria all'interno del processo, destinato a inverare le garanzie che circondano l'imputazione penale<sup>217</sup>.

Si conferma ancora una volta il ruolo cruciale del libero convincimento del giudice, che dovrebbe essere orientato alla luce del principio dell'oltre ogni ragionevole dubbio che, in quanto principio cardine del processo penale, deve abbracciare ogni elemento del reato, quindi sia la causalità che la colpevolezza. In altri termini, se anche le Sezioni Unite hanno recintato l'accertamento causale per impedire l'ingresso della teoria dell'aumento del rischio, nessuna breccia può essere creata in questo muro, neppure se essa è costruita per rispondere all'esigenza di tutela di beni giuridici, sebbene di rango elevato: l'elasticità dello schema del reato di evento sarebbe difficile da conciliare col rispetto dei principi fondamentali su cui si regge la costituzionalità dell'ordinamento penale, cioè il principio della responsabilità personale e il principio di legalità. Queste conclusioni confermano l'idea secondo la quale "il problema causale in diritto penale non avrebbe natura tecnica, bensì squisitamente politico-criminale" perché "il favore o il disfavore nei confronti di soluzioni garantiste dipende dall'universo assiologico dell'interprete, più che dalle contenzioni della "dogmatica"<sup>218</sup>.

L' "universo assiologico" cui ho fatto qui riferimento richiamando il principio di legalità e di responsabilità personale dovrebbe guidare non solo il lavoro dell'interprete, ma anche quello del legislatore. Se infatti uno stato di incertezza è incompatibile con il diritto penale d'evento, questo non annulla l'intervento del diritto penale come strumento di tutela dei beni giuridici, piuttosto dovrà orientarlo: così come il principio di legalità impone rigidità nell'accertamento della responsabilità, lo stesso si concilia con categorie più flessibili attraverso lo schema dei reati di pericolo. Un modello punitivo che guarda al futuro, quale è quello del pericolo astratto, è ritenuto più compatibile con il rischio, rispetto al diritto penale d'evento che ha lo sguardo rivolto al passato<sup>219</sup>. Ma attenzione:

---

<sup>217</sup> G. De Francesco, *L'imputazione del reato*, cit., pp. 536-537.

<sup>218</sup> O. Di Giovine, *op. cit.*, p. 210-211.

<sup>219</sup> C.E. Paliero, *Responsabilità per danno da prodotto tra persona fisica ed ente collettivo*, in *Il diritto penale di fronte alle sfide della "società del rischio"*, cit., p. 291.

non ogni rischio potrà essere penalmente rilevante, poiché è noto che l'idoneità della condotta a mettere in pericolo i beni tutelati deve poggiare su basi scientifiche solide, che presentino come ragionevole la risposta punitiva. In altri termini, la rigidità del diritto penale d'evento che si prospetta in generale nella letteratura sui danni da prodotto non può essere letta come accondiscendenza verso un illecito penale basato sul principio di precauzione. Inoltre, va evidenziato che l'eventuale futura elaborazione di fattispecie di pericolo astratto non comporterà una totale rinuncia *a priori* al diritto penale d'evento, che può e deve farsi spazio in un contesto di rischio. In questa direzione rappresenta una risposta l'affinamento di tecnologie in grado di leggere i meccanismi causali che governano la decisione algoritmica, come si è avuto modo di illustrare; o l'elaborazione di regole che consentano il tracciamento del processo produttivo dei manufatti intelligenti.

Ma i reati causalmente orientati non potranno essere l'unica risposta, poiché allo stato non pare possano occupare una posizione di primo piano nella tutela dei beni giuridici, a meno di sacrificare alcune conquiste della civiltà giuridica moderna. Per questo bisognerà valorizzare prospettive di politica legislativa atte ad offrire forme di reazione penali e extrapenali, come già in altri contesti di modernità era stato auspicato dalla dottrina<sup>220</sup>.

Qui si può aggiungere che una prospettiva da valorizzare è anche quella della responsabilizzazione dell'ente, già sostenuta in dottrina nell'ambito della più generale analisi dei danni da prodotto<sup>221</sup> e, da ultimo, anche nel contesto dei nuovi rischi creato dalla diffusione delle tecnologie intelligenti<sup>222</sup>.

Si tratta di risposte che la prassi solleciterà al diritto nel momento in cui ci si avvierà verso un utilizzo capillare dei sistemi intelligenti che forgerà nuove esperienze criminologiche alle quali il diritto cercherà di dare risposte sempre più adeguate.

---

<sup>220</sup> F. Stella, *Giustizia e modernità*, cit., pp. 221 ss.; C. Piergallini, *op. cit.*, p. 105 sulla flessibilità dei modelli di tutela civilistica e pp. 469 ss. per la costruzione della nuova categoria dell'illecito penale di rischio nella forma di reato d'obbligo a forma vincolata, categoria però criticata in dottrina, v. ad es. G. De Francesco, *L'imputazione del reato*, cit., pp.530 ss.; *Id.*, *Dinamiche del rischio e modelli d'incriminazione nel campo della circolazione di prodotti alimentari*, in *Per uno studio interdisciplinare su agricoltura e alimentazione*, M. Goldoni, E. Sirsi (a cura di), Milano, Giuffrè, 2010, pp. 298 ss.; in ambito medico, a favore di un ricorso a strumenti alternativi alla risposta penale anche P. Veneziani, *op. cit.*, p. 2000.

<sup>221</sup> F. D'Alessandro, *Rischio e responsabilità amministrativa degli enti*, in *Il diritto penale di fronte alle sfide della "società del rischio"*, cit., pp. 90-111; M.A. Bartolucci, *L'art. 8 d.lgs. n. 231 del 2001 nel triangolo di Penrose, tra minimizzazione del rischio-reato d'impresa e "nuove forme" di colpevolezza*, *ibidem*, pp. 112-133.

<sup>222</sup> F. Consulich, *Il nastro di Mobius. Intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato*, in *Banca, borsa, tit. credito*, 2018, pp. 224 ss.

### SEZIONE III

## LA COLPEVOLEZZA NEL REATO COLPOSO DEL PRODUTTORE E DELL'UTENTE DEL SISTEMA DI IA

1. Colpevolezza colposa – 2. L'accertamento della misura soggettiva della colpa – 3. La colpevolezza colposa del produttore – 4. La posizione dell'utente – 5. Le circostanze anormali tra la colpevolezza e il concetto di azione "cosciente e volontaria" – 6. Caso fortuito, forza maggiore e colpevolezza colposa – 7. L'inquadramento di alcuni fattori problematici: a) l'autoapprendimento – 7.1 (segue) b) Manomissione e manipolazione dell'algoritmo – 7.2 (segue) c) Cattiva informazione o informazione incompleta dell'utente – 7.3 (segue) d) Il malfunzionamento improvviso e incontrollabile

### 1. Colpevolezza colposa

Nell'impostazione del reato colposo qui seguita, l'analisi della responsabilità colposa del produttore e dell'utente di un sistema intelligente, dopo aver vagliato i punti critici attinenti al fatto tipico colposo e alla causalità, deve concentrarsi sulla colpevolezza colposa. Malgrado le tendenze oggettivizzanti che spesso si registrano nella giurisprudenza<sup>1</sup>, la dottrina è decisa nel sottolineare la centralità che nel giudizio penale deve avere la dimensione soggettiva della colpa<sup>2</sup>, dove si assiste al passaggio dalla "impersonalità" dell'agente modello alla "personalità dell'agente reale, attraendo nell'orbita dell'accertamento le sue capacità"<sup>3</sup>, così da soggettivizzare l'esigibilità della richiesta di osservare la regola cautelare<sup>4</sup>.

---

<sup>1</sup> Come ha cura di precisare S. Dovere, *Giurisprudenza della Corte suprema sulla colpa*, in *Enc Dir., I Tematici, Il reato colposo*, Milano, Giuffrè, p. 579 ss., p. 600, "per quanto sia ormai nitido il riconoscimento che la giurisprudenza tributa alla misura soggettiva della colpa, non si può ignorare che questo fattore essenziale del giudizio di responsabilità trova ancora uno spazio troppo angusto nell'accertamento giudiziale".

<sup>2</sup> D. Castronuovo, *La colpa penale*, Milano, Giuffrè, 2009; M. Donini, *Teoria del reato*, Padova, Cedam, 1996, pp. 348 ss.; *Id.*, *L'elemento soggettivo della colpa. garanzie e sistematica*, in *Riv. it. dir. proc. pen.*, 2013, pp. 124 – 156; S. Canestrari, *La doppia misura della colpa nella struttura del reato colposo*, in *Ind. pen.*, 2012, Pp. 21-30; M. Grotto, *Principio di colpevolezza, rimproverabilità soggettiva e colpa specifica*, Torino, Giappichelli, 2012; A. Canepa, *L'imputazione soggettiva della colpa. Il reato colposo come punto cruciale nel rapporto tra illecito e colpevolezza*, Torino, Giappichelli, 2011.

<sup>3</sup> C. Piergallini, voce *Colpa (diritto penale)*, *Enc. dir. – Annali X*, 2017, pp. 222 -265, in particolare p. 244.

<sup>4</sup> M. Donini, *L'elemento soggettivo*, *cit.*, p. 147.

La colpevolezza colposa gioca un ruolo fondamentale perché può escludere la rilevanza penale della condotta violativa della regola cautelare ovvero può intervenire sul grado della colpa e, quindi, in sede di commisurazione della pena<sup>5</sup>.

Questo accertamento si prefigge di analizzare la fattispecie concreta con la lente non più dell'agente modello, ma dell'agente concreto, del quale si tengono in considerazione aspetti maggiormente personali, ovvero aspetti relativi al caso concreto. Si deve allora accertare che l'imputato *poteva* attenersi alla regola cautelare, dopo aver verificato che *doveva* farlo alla luce del parametro dell'agente modello.

Certo il problema di questa impostazione è noto: posto che un giudizio totalmente personalizzato porta automaticamente a escludere la responsabilità<sup>6</sup> e che quindi è necessaria una certa astrazione dalle circostanze concrete<sup>7</sup>, per cui “anche la misura soggettiva corrispond[e] in fondo a un punto di vista oggettivo, in quanto anche essa contiene quel margine di generalizzazione indispensabile per poter operare il giudizio di colpa”<sup>8</sup>, ci si è domandato quali caratteristiche dovessero essere prese in considerazione dal giudice per la valutazione della colpevolezza. In cosa consiste allora la misura soggettiva della colpa?

La conclusione secondo la quale integrano la misura soggettiva della colpa da un lato, l'elemento delle particolari caratteristiche individuali del soggetto agente e, dall'altro, l'esigibilità tesa a connotare l'eccezionalità di una situazione di fatto che condiziona l'effettiva capacità al rispetto della regola cautelare<sup>9</sup>, va qui condivisa.

Prima di calarci ad analizzare questi due aspetti con riguardo alle figure del produttore e dell'utente di un sistema intelligente, è doveroso sviluppare alcune considerazioni preliminari, al fine di riassumere gli approdi della dottrina sulla colpevolezza colposa, non mancando di notare sin da ora come scarso sia invece

---

<sup>5</sup> In quest'ultima prospettiva, D. Castronuovo, *La colpa "penale". Misura soggettiva e colpa grave*, in *Riv. it. dir. proc. pen.* 2013, pp. 1723 ss. Sull'esigenza di estendere l'imputazione ai soli casi di colpa grave, M. Donini, *L'elemento soggettivo*, cit., p. 154-156, il quale evidenzia che la colpa lieve sul piano probatorio tende a coincidere con la violazione di una cautela impersonale, attestandosi sulla mera irregolarità e rischio non consentito, rendendosi espressione della prevenzione generale e della funzione politico-sociologica della pena che contrasta con il principio di colpevolezza.

<sup>6</sup> G. Marinucci, *La colpa per inosservanza di leggi*, cit., p. 183.

<sup>7</sup> Astrazione che, come precisa V. De Francesco, è il punto che accomuna le tesi oggettiviste e soggettiviste, poiché, sebbene gli ultimi si mostrino favorevoli a una maggiore individualizzazione, sono comunque consapevoli del fatto che occorre fare astrazione almeno da uno degli elementi del fatto concreto. V. De Francesco, *Sulla misura soggettiva della colpa*. *Studi Urbinati, A, Scienze giuridiche, politiche ed economiche*, 30., pp. 273-343, in particolare p. 295.

<sup>8</sup> V. De Francesco, cit., p. 303.

<sup>9</sup> G. Fornasari, *Il principio di inesigibilità nel diritto penale*, Padova, CEDAM, 1990, p. 327 e 334.

l'interesse della giurisprudenza a volgere lo sguardo agli aspetti più soggettivi della colpa una volta accertata la violazione della regola cautelare alla luce del parametro dell'agente modello<sup>10</sup>.

Quanto al primo di questi due elementi, cioè le caratteristiche individuali del soggetto agente, nel dibattito registratosi tra “oggettivisti” e “soggettivisti”<sup>11</sup> relativo alle circostanze concrete dalle quali bisogna fare astrazione per operare il rimprovero di colpa<sup>12</sup>, prevale nella dottrina italiana l'opinione di quanti sostengono che le caratteristiche personali dell'agente rilevanti in sede di colpevolezza sono quelle intellettuali e fisiche, quindi solo i limiti attinenti a tali aspetti avranno un peso in sede di accertamento processuale<sup>13</sup>. Parimenti non potrà dirsi per i tratti caratteriali, i quali saranno invece irrilevanti. Assegnare un ruolo in sede di accertamento della colpa alle caratteristiche fisiche e intellettuali significa dare un peso maggiore al principio di colpevolezza piuttosto che alle esigenze di prevenzione generale, in linea con l'idea secondo la quale la colpevolezza non è un carattere negoziabile del modello costituzionale di reato, neppure di quello colposo<sup>14</sup>.

---

<sup>10</sup> La scarsa attenzione della giurisprudenza alla tematica della colpevolezza colposa è messa in evidenza da tutti gli autori citati *supra*, nota 1. È però vero, come ben evidenziano gli stessi autori, che il giudizio non è assoluto, perché il problema è stato affrontato talvolta implicitamente (come nel caso *Azzini*, *infra*, in nota, del 1997), talaltra esplicitamente, come in Cass. sez. un., n. 38343/2014 (*Thyssenkrupp*) e Cass. sez. IV, n. 12175/2017. Per un'analisi della giurisprudenza italiana e tedesca, si rinvia a Canepa, *cit.*, cap. IV, pp. 109 ss.

<sup>11</sup> Su cui si rinvia a V. De Francesco, *cit.*.

<sup>12</sup> È proprio su questo punto che verterebbe il contrasto tra le due posizioni dottrinali, secondo V. De Francesco, *cit.*, p. 296. Infatti, precisa l'A., che tutti i soggettivisti sono d'accordo sulla necessità di prescindere da almeno una delle condizioni necessarie per il verificarsi del fatto colposo, poiché diversamente l'esito sarebbe sempre assolutorio.

<sup>13</sup> G. Fiandaca, E. Musco, *Diritto penale, p.te gen.*, Torino, Giappichelli, 2019, p. 607; C. Piergallini, voce *Colpa*, *cit.*, p. 244; S. Canestrari, *La doppia misura*, *cit.*, p. 29; M. Romano, *Commentario sistematico al codice penale, art. 43*, p. 468. Diversa l'opinione di G. Marinucci, *Il reato come “azione”*, Milano, Giuffrè, 1970, p. 209 ss., il quale ritiene che non è possibile calare il giudizio della misura subiettiva nelle caratteristiche dell'agente in concreto essendo questa un'operazione *logicamente impossibile*, posto che la “misura di giudizio” deve trovarsi “al di fuori” dell'oggetto da misurare e non può di certo coincidere con esso. Per cui non vi è misura soggettiva di giudizio quando l'oggetto da misurare (cioè le capacità individuali dell'agente emerse nell'azione concreta) costituiscono anche la misura che si adotta. Nondimeno, anche Marinucci ritiene che il giudizio sulla colpa non possa arrestarsi alla constatata violazione di una regola di diligenza, ma limita l'ulteriore accertamento alla sola presenza di circostanze anormali nel momento in cui si è tenuta la condotta negligente, al fine di comprendere se esse hanno inciso sull'azione negligente o l'abbiano financo determinata.

<sup>14</sup> D. Castronuovo, voce *Colpa penale*, *Enc. dir., I Tematici, Il reato colposo*, Milano, Giuffrè, pp. 200 ss., a p. 228 e 230.

## 1.1 L'esigibilità

Più complesso il discorso sviluppatosi intorno al concetto di esigibilità.

Si tratta di una categoria o principio (solitamente declinato in negativo come inesigibilità) che rileva già in altri momenti del reato, avendo un ruolo nella definizione dei doveri che fondano il carattere tipico e antigiuridico di un comportamento umano<sup>15</sup> e che, in ultima istanza, ha un ruolo in sede di colpevolezza del reato, tanto doloso, quanto colposo.

La dottrina infatti ha studiato *l'inesigibilità* come fondamento di scusanti comuni o previste da specifiche norme<sup>16</sup>, per volgere poi l'attenzione all'*esigibilità* come possibile criterio di individuazione dell'evitabilità o inevitabilità (e quindi della scusabilità) dell'errore sul precetto penale e, infine, come criterio regolativo della misura soggettiva del dovere di diligenza<sup>17</sup>. Si conclude che è proprio la colpa la "sede privilegiata del discorso sull'inesigibilità", al pari dell'omissione, in ragione della natura normativa della colpa<sup>18</sup> e della necessità di tener conto del lato anche soggettivo del dovere di diligenza dopo averne accertato quello oggettivo che connota l'illiceità colposa. L'inesigibilità si presenterebbe allora come criterio di sussistenza e graduazione della colpa<sup>19</sup>.

---

<sup>15</sup> Sul ruolo dell'inesigibilità anche in sede di tipicità e di antigiuridicità del fatto, si rinvia alla già citata opera di G. Fornasari. In giurisprudenza, nota M. Grotto, *Principio di colpevolezza, rimproverabilità soggettiva e colpa specifica*, Torino, Giappichelli, 2012, p. 334, che una pronuncia del 1990 definisce l'inesigibilità alternativamente come "causa di esclusione dell'antigiuridicità del fatto o di esclusione della colpevolezza" (Cass. 28 nov. 1990, in *Riv. pen.*, 1991, 945 e in *Riv. trim. dir. pen. ec.*, 1991, p. 813. In senso critico verso l'accoglimento della categoria dell'inesigibilità, F. Mantovani, *Diritto penale, cit.*, p. 360, il quale mette in evidenza come infondata sia l'esigenza di sistemare sotto un *genus* comune una serie di scusanti codificate, poiché esse sarebbero inquadrabili tra le scriminanti (stato di necessità e art. 384 c.p.) o le cause di esclusione del nesso di causalità (caso fortuito e forza maggiore) e come configurare l'inesigibilità come causa generale di esclusione della colpevolezza comporti un grave pregiudizio della certezza del diritto e quindi la violazione del principio di legalità.

<sup>16</sup> È questo il caso della previsione dell'art. 384 c.1 c.p., la cui natura giuridica è stata a lungo dibattuta in dottrina e in giurisprudenza tra quella di scriminante speciale di stato di necessità e di scusante. Di recente le Sez. Un. della Cassazione (sent. 26 novembre 2020 (dep. 17 marzo 2021), n. 10381, Pres. Cassano, rel. Fidelbo, ric. Fiavola) hanno aderito a quest'ultima impostazione, già sostenuta da tempo da una parte della dottrina, tra cui lo stesso Fornasari, *op. ult. cit.*, p. 353 ss., precisando che "in quanto causa di esclusione della colpevolezza, è applicabile analogicamente anche a chi ha commesso uno dei reati ivi indicati per esservi stato costretto dalla necessità di salvare il convivente more uxorio da un grave e inevitabile nocumento nella libertà e nell'onore".

<sup>17</sup> G. Fornasari, *cit.*, pp. 320 ss.

<sup>18</sup> *Ibidem*, p. 320.

<sup>19</sup> D. Castronuovo, *La colpa penale*, Milano, Giuffrè, 2009, p. 564. In questo studio si è però già fatto ricorso al concetto di inesigibilità per escludere o limitare il dovere di diligenza nel momento oggettivo della colpa, definendo inesigibile sul piano del *dover* essere la richiesta, rivolta al produttore, di sviluppare un algoritmo che conosca sempre tutte le circostanze del caso concreto. Valga quindi una precisazione. Chi sostiene la rilevanza dell'esigibilità in sede di colpevolezza, tende a precisare che nessun ruolo possa avere già nel momento dell'illiceità, criticando le impostazioni che leggono il dovere di diligenza che connota il fatto tipico alla luce del principio di inesigibilità. Si dice, in particolare, "delle due l'una: o si ritiene

Quanto al contenuto che la categoria dell'esigibilità ha in sede di colpevolezza, si ritiene che essa prenda in considerazione situazioni fattuali eccezionali, non standardizzabili, poiché diversamente già rilevanti in sede di illiceità colposa nel delineare la figura differenziata di agente modello<sup>20</sup>. In particolare, accade che la situazione presenti aspetti non generalizzabili e, per questo, non rilevabili al momento della formulazione della corrispondente figura dell'agente modello<sup>21</sup>; nondimeno, l'accertamento della penale responsabilità non può mostrarsi insensibile a tali vicende peculiari, che, per le loro caratteristiche, si rivelano invece decisive nell'escludere o diminuire il rimprovero. Il soggetto è quindi perfettamente imputabile, ma versa in una condizione di incapacità o diminuita capacità situazionale, momentanea, contingente, incorrendo in un errore esecutivo (che incide sul *fare*) o in un errore cognitivo o di rappresentazione (che influenza il sapere ontologico o nomologico), di cui bisogna tener conto per valutare se sia scusabile o quantomeno rilevante sul piano commisurativo<sup>22</sup>.

È possibile che tali fattori situazioni incidano sul soggetto procurando al suo interno un conflitto motivazionale; nondimeno, delle volte tale conflitto può mancare, eppure è comunque non esigibile la richiesta rivolta all'agente concreto di conformare il proprio comportamento al parametro cautelare, in ragione di una particolare difficoltà dovuta a circostanze che incidono concretamente sulla riconoscibilità del rischio, sulla prevedibilità dell'evento o sull'effettiva possibilità di mantenersi nell'area del rischio lecito o di eseguire correttamente l'attività intrapresa<sup>23</sup>.

---

veramente che ogni fattore soggettivo e individuale incida sull'esistenza del dovere oggettivamente inteso" cosicché già la misura oggettiva sarebbe relativizzata, "oppure si punta a mantenere la distinzione tra un aspetto oggettivo e un aspetto soggettivo del dovere di diligenza (...) e in tal caso per il primo di essi non può che valere una misura generalizzata (...) e pertanto non influenzabile dal criterio dell'esigibilità" (G. Fornasari, *cit.*, p. 331). L'obiezione è condivisibile, ma è anche vero che in sede di *Unrecht* possono valere dei fattori generali e non individuali che limitano già verso il basso il dovere di diligenza, rendendo un livello maggiore inesigibile. Così accade nel caso del nostro produttore, al quale non è possibile già in sede oggettiva richiedere di produrre un *software* che conosca tutte le situazioni possibili in cui sarà impiegato: è, questa, una condizione che vale per qualsiasi produttore e non per uno solo o per una certa categoria, perché l'esperienza futura non è trasmissibile visto non si è ancora verificata! Per questo vi è una porzione di diligenza che non può essere richiesta neanche all'agente modello che allora diventa inesigibile già sul piano oggettivo, mancando un dovere oggettivamente esistente.

<sup>20</sup> D. Castronuovo, *La colpa penale, cit.*, p. 565., il quale porta come esempi i casi del conducente principiante o all'automobilista disabile o alla figura del medico specializzando. *Id.* D. Castronuovo, *La colpa "penale". Misura soggettiva e colpa grave*, in *Riv. it. dir. proc. pen.*, 2013, pp. 1723 ss., in particolare p. 1752.

<sup>21</sup> D. Castronuovo, *La colpa penale, cit.*, p. 564.

<sup>22</sup> D. Castronuovo, *La colpa, cit.*, p. 565 e p. 591.

<sup>23</sup> D. Castronuovo, *La colpa penale, cit.*, p. 599, il quale riassume le indicazioni provenienti dalla giurisprudenza sulle ipotesi individualizzanti del giudizio di colpa, elaborando un catalogo che non si limita a comprendere le sole ipotesi di conflitto motivazionale.

Qualcuno limita il concetto di esigibilità ai soli casi di conflitto motivazionale tra il dovere scaturente dalla norma e un ostacolo esterno che vi si frappone<sup>24</sup>. Sul punto si può osservare che è proprio in questi termini che la categoria dell'esigibilità è stata positivizzata, come dimostrano, a titolo di esempio, lo stato di necessità scusante di cui all'art. 54 c.p. e la causa di non punibilità di cui all'art. 384 c.p.<sup>25</sup>, entrambe pronte a dare spazio al conflitto tra valori che ha determinato il soggetto a commettere l'illecito. Ma se questa è la scelta del legislatore, ridurre al conflitto motivazionale anche la dimensione non positivizzata del principio di inesigibilità significa forse costringere l'interprete a una lettura troppo restrittiva dello stesso. Piuttosto, le particolari difficoltà del caso concreto dovrebbero rilevare proprio in tema di reato colposo, dove l'inesigibilità può trovare spazio anche quando un conflitto motivazionale non ci sia, ma sussista una impossibilità concreta di allinearsi al dovere cautelare, non standardizzabile e quindi non incidente a livello di illiceità colposa.

Occorre a questo punto fare una precisazione linguistica, che ha però importanti ricadute sul piano sostanziale e, soprattutto, processuale. Quando ci occupiamo di definire queste situazioni che determinano una incapacità o diminuita capacità temporanea e contingente, riconosciamo loro una valenza negativa nella colpevolezza, nel senso che si tratta di situazioni che non devono ricorrere affinché la colpa sussista. Rovesciando i termini del discorso, diremmo che la colpevolezza colposa si compone dell'esigibilità, la cui definizione sarà ricavata in negativo come assenza di condizioni contingenti che determinano un errore esecutivo o di rappresentazione del soggetto agente in grado di escludere la colpevolezza. Così, mentre l'inesigibilità esclude la colpevolezza, l'esigibilità la fonda<sup>26</sup> e valorizzare l'esigibilità come elemento strutturale della colpevolezza colposa comporta che il relativo onere probatorio grava sulla pubblica accusa, mentre a definirla in termini negativi, cioè come inesigibilità, essa graverebbe sulla difesa<sup>27</sup>.

---

<sup>24</sup> Così A. Canepa, *cit.*, p. 206-208, il quale ritiene che il fattore esterno che incida sulla gravosità dello sforzo diligente rilevarebbe però in sede oggettiva, di *Unrecht*, perché il fatto è reso inevitabile dall'agente, così escludendo il carattere illecito della condotta.

<sup>25</sup> Sulla cui natura di scusante, da ultimo Cass., Sez. un., 16 marzo 2021, n. 10381, che ha accolto la tesi dell'estensione della stessa anche ai conviventi *more uxorio*.

<sup>26</sup> M. Grotto, *Principio di colpevolezza, cit.*, p. 349.

<sup>27</sup> M. Grotto, *Principio di colpevolezza, cit.*, p. 351; al recupero in termini positivi della categoria dell'esigibilità è poi dedicato l'intero cap. III, pp. 355 ss. in cui si precisa che il giudizio di colpevolezza è duplice, nel senso che occorre prima procedere a fondare la colpevolezza utilizzando i fattori soggettivi emergenti dal caso concreto per sussumerlo entro una norma astratta e, successivamente, occorre verificare che non vi fossero speciali condizioni di fatto tali da elidere la responsabilità penale sul piano della

Abbiamo quindi individuato le ragioni sostanziali e processuali per valorizzare l'esigibilità come elemento che fonda la colpa: riempire la categoria della colpevolezza colposa individuandone gli elementi per indirizzare l'onere probatorio dell'accusa in sede processuale e quello motivazionale del giudice, al fine di stabilire se l'agente aveva il dovere e il potere di adeguarsi al paradigma cautelare.

In sostanza, l'oggetto dell'accertamento di colpevolezza si colora ancora una volta attraverso la prevedibilità o riconoscibilità dell'evento o della situazione di pericolo e l'evitabilità degli stessi, nondimeno il parametro è qui soggettivo, dovendosi avere riguardo alle possibilità dell'agente concreto, quindi a circostanze suscettibili di incidere concretamente sulla prevedibilità e evitabilità<sup>28</sup>. La differenza è sottolineata in dottrina precisando che nella misura soggettiva la riconoscibilità e l'evitabilità vengono in rilievo come *regola di giudizio* e non più come regola di condotta<sup>29</sup>. In particolare, si osserva che la misura soggettiva rileva a livello fattuale, cioè di tipicità soggettiva, per definire la regola di condotta, ricostruendola e generalizzandola in base ai tipi di soggetti e alle situazioni; a livello di colpevolezza, invece, la misura soggettiva confluisce in una regola di giudizio, cioè in una scusante o in "un limite normativo di applicazione di regole di condotta, che attiene al profilo motivazionale o personale dell'esigibilità della condotta normativamente doverosa"<sup>30</sup>.

Tali regole di giudizio sono in parte codificate come scusanti, per altra parte invece non sono positivizzate, cosicché costituiscono "limiti logico-normativi all'esigibilità delle regole di condotta, ed attuazione della colpevolezza come *principio* giuridico, anziché mera espressione della colpevolezza come *categoria* dell'analisi del reato", che il giudice è chiamato ad attuare "in funzione di limite alle regole generalpreventive degli agenti-modello"<sup>31</sup>.

È però vero che anche con riguardo agli elementi situazionali si presenta l'esigenza di prenderne in considerazione solo alcuni, effettuando una selezione, così come accade

---

colpevolezza. Così, afferma l'A., "la [seconda] valutazione in termini di inesigibilità serve a correggere la già effettuata valutazione in termini di esigibilità", p. 356. Lo stesso A. in altro scritto precisa però che, in accordo con la dottrina maggioritaria, non sempre sarà necessaria la prova in positivo della colpevolezza colposa, essendo "del tutto fisiologico che, nella maggioranza delle ipotesi, il tema della sussistenza o meno della rimproverabilità soggettiva (e, dunque, il tema della colpevolezza colposa) non necessiti di essere affrontato nel corso dell'accertamento processuale. M. Grotto, voce *Scusanti della colpa*, in *Enc. dir., I Tematici, Reato colposo*, Giuffrè, Milano, 2021, pp 1153 ss., p. 1161.

<sup>28</sup> D. Castronuovo, *La colpa "penale"*, cit., p. 1751.

<sup>29</sup> D. Castronuovo, *La colpa "penale"*, cit., p. 1731.

<sup>30</sup> M. Donini, *L'elemento soggettivo*, cit., p. 149.

<sup>31</sup> M. Donini, *L'elemento soggettivo*, cit., p. 153.

con le caratteristiche proprie del soggetto agente. Pertanto, allo stesso modo in cui per le caratteristiche personali del soggetto agente si registra - come ricordato sopra - un certo accordo sulla non rilevanza di fattori caratteriali, anche per gli elementi situazionali bisogna interrogarsi su quali di essi tener conto al momento dell'individualizzazione del giudizio sulla colpevolezza colposa<sup>32</sup>.

La dottrina si è sforzata di indicare esempi più "aggiornati" rispetto a quelli proposti dalla manualistica che fanno riferimento a casi di scuola (anziane signore alle prese con le dinamiche della vita di città<sup>33</sup>). Il risultato è un catalogo di ipotesi esemplificative, che vale la pena di riportare qui, almeno nei suoi aspetti essenziali<sup>34</sup>. Sono quindi da valutare come fattori escludenti la colpevolezza colposa: a) inesperienza o minore capacità non rilevante in sede di tipicità soggettiva perché non standardizzabile, quindi incapace di escludere la violazione del dovere di diligenza; b) deficit culturale e sociale; c) stanchezza o debolezza improvvise; d) spavento, concitazione, sorpresa o disorientamento; e) speciale difficoltà e urgenza dell'intervento in situazione di imprevedibilità di un rischio specifico e di estrema agitazione e pressione psicologica<sup>35</sup>; f) perdita repentina di capacità funzionale che determina un errore esecutivo; g) conflitto apparente di doveri cautelari<sup>36</sup>; h) breve lasso di tempo a disposizione affinché la persona fisica si adegui alla regola di condotta ovvero l'ente adegui il modello organizzativo.

Se volessimo procedere a una ulteriore catalogazione delle ipotesi che la dottrina ha così individuato, potremmo riferire quelle *sub* a) e b) alle caratteristiche personali dell'agente, mentre le restanti ipotesi sarebbero raggruppabili entro la più generale categoria, anch'essa richiamata dalla dottrina, di "circostanze anormali", interne e esterne all'agente<sup>37</sup>, in grado di incidere sull'azione negligente o persino di determinarla.

---

<sup>32</sup> D. Castronuovo, *La colpa "penale"*, cit., p. 1730.

<sup>33</sup> Il riferimento è al noto esempio tratto da Schimdhäuser, *Strafrecht, A.T., Leherbuch*, 1974, p. 444, ripreso da V. De Francesco, *Sulla misura soggettiva della colpa. Studi Urbinati, A, Scienze giuridiche, politiche ed economiche*, 30., pp. 273-343, in particolare a p. 278.

<sup>34</sup> È il catalogo proposto da D. Castronuovo, *La colpa "penale"*, cit., pp. 1752 ss., che viene però qui riportato secondo un ordine diverso da quello col quale è presentato dall'autore; l'enunciazione qui proposta ci tornerà utile a breve.

<sup>35</sup> È l'ipotesi ritagliata sul caso della c.d. cervice uterina deciso da Cass., sez. IV, ed. 29 sett. 1997, n. 1693, Azzini, in *Riv. pen.*, 1998, p. 350 ss. alla quale ampio spazio è dedicato dalla trattazione di D. Castronuovo in *La colpa "penale"*, cit., pp. 1740 ss.

<sup>36</sup> Conflitto presente nel caso degli agenti della Polstrada, risolto da Cass., sez. IV, ud. 3 febr. 2009, n. 274, rel. Blaiotta, però escludendo la tipicità colposa, a dimostrazione della tendenza oggettivizzante con la quale la giurisprudenza si appropria al tema della colpa. Per un commento, D. Castronuovo, *La colpa "penale"*, cit., p. 1748 ss.

<sup>37</sup> G. Marinucci, *Il reato come "azione"*, Milano, Giuffrè, 1970, p. 217.

La dottrina ha ricondotto entro la categoria di “circostanze interne all’agente” i riflessi fisici dovuti a “momenti di terrore” (ad esempio l’invasione della corsia opposta a quella di marcia a causa dell’ingresso, nell’abitacolo della vettura, di uno sciame d’api) e i movimenti tenuti in stato di incoscienza (ad esempio la violazione della regola diligente a causa di un sopravvenuto malore alla guida), mentre sono classificate come “circostanze esterne” i movimenti causati per costringimento fisico e per forza maggiore<sup>38</sup>. Ma che tali situazioni, esterne o interne all’agente, rilevino in sede di colpevolezza non è da tutti condiviso; sul piano strutturale del reato, trattandosi di circostanze che incidono sulla dominabilità della condotta da parte dell’agente, esse possono rilevare quali precondizioni dell’azione, come si vedrà più avanti.

## 2. L’accertamento della misura soggettiva della colpa

Il catalogo sopra riportato ci sarà utile per calarci sulla scusabilità di reati verificatisi in occasione della produzione o dell’utilizzo di sistemi di IA. Procederemo allora secondo lo schema già tracciato in sede di analisi della misura oggettiva della colpa, affrontando separatamente le posizioni dell’utente e del produttore.

Prima però bisogna che ci domandiamo come va condotto l’accertamento sulla misura soggettiva della colpa. E infatti, dopo averne individuato le componenti nelle caratteristiche individuali del soggetto agente e nei fattori situazionali esterni, bisogna domandarsi in che modo il giudice dovrà condurre questo giudizio per escludere ovvero affermare la colpevolezza colposa.

Una corrente dottrinale richiama in questa sede il parametro dell’*homo eiusdem conditionis et professionis*<sup>39</sup>, di cui non si avvale in sede di fatto illecito, ricorrendo lì invece al parametro della miglior scienza e esperienza. In questo modo in sede di illiceità colposa si perviene a un giudizio identico per ogni agente, poiché non rileva

---

<sup>38</sup> *Ibidem*, p. 217 ss.

<sup>39</sup> F. Mantovani, *Diritto penale, p.te gen.*, Milano, Wolters Kluwer, 2017, p. 340; M. Grotto, *Principio di colpevolezza*, cit., p. 389 ss. In senso critico verso l’impostazione “iperoggettiva” di questo orientamento, M. Donini, *L’elemento soggettivo*, cit., p.141-142, nota 33, che commenta l’impostazione sposata da Grotto osservando, tra le altre obiezioni mosse, che essa “porta a identificare, sostanzialmente, la colpa oggettiva con il rischio consentito, per affrontare in chiave di colpevolezza la colpa soggettiva”.

l'appartenenza a una certa categoria sociale o professionale, elemento, questo, cui sarà data attenzione in sede di misura soggettiva<sup>40</sup>.

Secondo l'impostazione proposta da altra parte della dottrina, già seguita nel trattare le peculiarità del fatto tipico colposo nel contesto delle macchine intelligenti (SEZ. I di questo capitolo), il parametro dell'agente modello servirebbe a individualizzare il dovere di diligenza in sede di fatto tipico appunto, cosicché lo stesso criterio non potrebbe essere riproposto in sede soggettiva. Nondimeno è pur vero, come peraltro già notato, che un livello di astrazione è imprescindibile anche nella misura soggettiva, nel senso che non si potrà tener conto di tutti gli elementi del fatto così come concretamente verificatosi, che farebbero inevitabilmente concludere per un giudizio individuale di imprevedibilità o inevitabilità dell'evento: il giudice, allora, dovrà tener conto delle maggiori o minori conoscenze e capacità individuali dell'agente e delle eventuali circostanze fattuali peculiari che rendono inesigibile il comportamento conforme a diligenza, "chiedendosi se un altro soggetto dotato delle conoscenze, energie, esperienze dell'agente al momento del fatto (ovvero l'agente stesso "pensato" senza la condotta in questione) sarebbe secondo la comune esperienza stato in grado di osservare la regola di diligenza violata"<sup>41</sup>. Questo procedimento di astrazione è da taluni descritto affermando che, ancora una volta, si seleziona "una figura-modello, diversa, però, da quella rilevante sul terreno del confezionamento del dovere oggettivo di diligenza", che prenda in considerazione "solo alcune delle caratteristiche fisiche e intellettuali dell'agente concreto"<sup>42</sup>.

Che quindi questo processo di astrazione dicasi effettuato ricorrendo a un agente modello, non significa equiparare la misura oggettiva a quella soggettiva, poiché si tratta di una figura modello diversa, nonostante "la misura del potere dell'agente concreto è pur sempre il potere di un altro"<sup>43</sup> e il giudizio è "strutturalmente identico, anzi, a quello della misura oggettiva della colpa" ed è anch'esso "giudizio di probabilità" che individualizza il rimprovero "secondo le esigenze proprie della colpevolezza"<sup>44</sup>.

---

<sup>40</sup> Quindi se il ragioniere di banca si improvvisa nel fine-settimana elettricista, sarà giudicato secondo l'agente-modello non del bancario, ma dell'elettricista, in base al relativo livello di perizia, così F. Mantovani, *op. ult. cit.*, p. 340.

<sup>41</sup> M. Romano, *Commentario, cit.*, p. 468.

<sup>42</sup> C. Piergallini, *voce Colpa, cit.*, p. 244.

<sup>43</sup> Così V. De Francesco, *Sulla misura, cit.*, p. 305, riprendendo le affermazioni di A. Kaufmann, *Das Schuldprinzip*, 1961.

<sup>44</sup> M. Romano, *Commentario, cit.*, p. 468.

In conclusione, anche la misura soggettiva corrisponde a un punto di vista oggettivo, atteso che in essa si riflette quel margine di generalizzazione indispensabile per poter operare il giudizio di colpa<sup>45</sup>.

### **3. La colpevolezza colposa del produttore**

Accogliendo l'impostazione della dottrina che afferma che a delimitare il dovere di diligenza nel suo aspetto soggettivo concorrono le caratteristiche individuali e l'esigibilità, quest'ultima concernente "l'eccezionalità di una situazione di fatto che condiziona l'effettiva capacità al rispetto della regola cautelare" "e non aspetti della personalità dell'agente"<sup>46</sup>, va analizzata più da vicino la posizione del produttore.

a) Riflettendo sulle caratteristiche individuali, potrebbero venire in rilievo le dimensioni dell'impresa. Aderendo però all'idea che debba trattarsi di caratteri non standardizzabili, in quanto altrimenti rilevanti ai fini della definizione di una figura differenziata di agente modello<sup>47</sup>, sembra più corretto ritenere che le dimensioni dell'impresa rilevino sul piano del dovere impersonale di conformarsi al parametro cautelare. Si tratterebbe, a ben vedere, di condizioni non uniche, ma che caratterizzano ogni situazione produttiva; inoltre, valga a tal proposito ricordare che le dimensioni dell'impresa hanno già rilievo a livello legislativo, nazionale e europeo, per taluni fini e, per tale ragione, sono oggetto di una specifica classificazione<sup>48</sup>. Conseguentemente, le dimensioni dell'impresa sono bensì rilevanti, ma nella definizione di agenti modello differenziati, quindi in punto di fatto tipico, cioè di dovere di riconoscere e evitare il rischio o l'evento<sup>49</sup>.

b) Anche le ipotesi di eccezionalità della situazione di fatto che condiziona la riconoscibilità del rischio e il rispetto della regola cautelare possono connotare l'accertamento della colpevolezza colposa.

---

<sup>45</sup> V. De Francesco, *op. cit.*, p. 303.

<sup>46</sup> È questa la ricostruzione di G. Fornasari, *op. cit.*, pp. 327 e 334, di cui si è dato atto nel paragrafo precedente.

<sup>47</sup> Così D. Castronuovo, *La colpa penale, cit.*, p. 565, già richiamato *supra*.

<sup>48</sup> SI fa riferimento alla raccomandazione 2003/361/CE della Commissione, relativa alla definizione delle microimprese, piccole e medie imprese, nel cui allegato è contenuta la definizione di microimpresa, piccola e media impresa (PMI) (art. 2).

<sup>49</sup> Si rinvia a quanto osservato nella sez. I di questo capitolo, al par. 6.2.2.

b.1) Un effetto scusante potrà avere, ad esempio, l'evoluzione imprevedibile dell'algoritmo a seguito dell'autoapprendimento, là dove l'imprevedibilità non sia quella che caratterizza ogni sistema di *machine learning* - che in quanto tale rilevarebbe, come si è tentato di dimostrare nella SEZ. I, già in sede di individuazione della regola di diligenza – ma sia dovuta a circostanze eccezionali antecedenti o concomitanti rispetto all'utilizzo del sistema. Si pensi, ad esempio, al sistema che è dotato di un autoapprendimento controllato, che blocchi possibili sviluppi degenerativi dell'apprendimento, inviando dei segnali alla casa produttrice. Se il sistema è manomesso dall'utente o anche da un terzo, come ad esempio un *hacker*, i blocchi impostati dal produttore saranno superati e l'autoapprendimento non avverrà più secondo quegli standard indicati dal produttore<sup>50</sup>. Laddove si verifici un fatto penalmente rilevante, l'eccezionalità della situazione potrà essere valutata per escludere il rimprovero per colpa, salva sempre la responsabilità del produttore allorché si dimostri che il sistema non era dotato di adeguati meccanismi di sicurezza per proteggerlo da attacchi esterni.

b.2) Un'altra circostanza di cui tener conto per verificare la possibilità dell'agente concreto di attenersi al dettato cautelare è rappresentata dall'obsolescenza della regola cautelare a seguito della scoperta di una nuova procedura di produzione che garantisce una maggiore sicurezza del prodotto. Secondo il parametro dell'agente modello, "corretto" alla luce di un'impostazione deontologica, vanno adottate le misure preventive tecnologicamente più evolute ed efficaci, quindi ogni azienda sarà tenuta a rispettarle, diversamente il suo comportamento sarà valutato come negligente già sul piano impersonale del dovere di diligenza<sup>51</sup>. Tuttavia, cosa accade se il nuovo standard produttivo scoperto da una certa azienda è coperto dal segreto industriale?<sup>52</sup> In questo caso, solo l'autore della scoperta è in grado di attenersi al nuovo dettato cautelare, mentre una simile possibilità non vi sarà per tutti gli altri produttori, che non sono messi nelle condizioni di conoscere la regola. La questione è se tale circostanza, cioè l'impossibilità di conoscere la regola coperta da segreto, rilevi sul piano della tipicità per calibrare l'aspettativa ordinamentale nei confronti dell'agente modello, ovvero sul piano soggettivo della possibilità di adeguamento alla regola cautelare.

---

<sup>50</sup> Nota B. W., Smith, *Automated Driving and Product Liability*, *Michigan State Law Review* 2017, no. 1 (2017), pp. 1-74, p. 51, come tali vulnerabilità siano dovute non tanto all'automazione, quanto piuttosto alla circostanza che i nuovi veicoli siano connessi alla rete, cosicché sorgono questioni inedite nel contesto dei sinistri stradali, legate alla *cybersecurity vulnerability*.

<sup>51</sup> G. Fiandaca, E. Musco, *Diritto penale*, cit., p. 586.

<sup>52</sup> Riflette sull'argomento M. Grotto, *Principio di colpevolezza*, cit., p. 375.

A voler aderire alla tesi che ricostruisce la tipicità colposa in modo estremamente oggettivo attraverso il parametro della miglior scienza ed esperienza, per poi dar spazio solo in sede di colpevolezza alle situazioni individualizzanti di cui tener conto alla luce del parametro dell'*homo eiusdem condicionis et professionis*<sup>53</sup>, la conclusione sarebbe la seguente: la tipicità è integrata perché la regola cautelare è già disponibile, nondimeno il rimprovero per colpa verrà meno in sede di colpevolezza, in ragione dell'impossibilità di conoscere la predetta regola in quanto coperta da segreto; ciononostante, la vittima avrà diritto al risarcimento secondo le norme civilistiche, sussistendo la violazione della regola cautelare<sup>54</sup>.

Nel presente lavoro si è invece sostenuta una tesi "mediana", che valorizza l'individualizzazione del giudizio di colpa già in sede di fatto tipico per il tramite del parametro dell'*homo eiusdem condicionis et professionis*, ma adeguato alla luce del modello deontologico, a mente del quale l'agente modello dovrà comunque agire secondo la migliore scienza ed esperienza<sup>55</sup>. In sostanza, non si rinuncia del tutto a individualizzare il giudizio sul dovere di diligenza, pur richiedendo all'agente modello di adeguarsi non agli usi sociali diffusi, ma agli standard più elevati disponibili a livello scientifico-tecnologico in quel momento.

Ebbene, in base a questo parametro, è vero che la regola cautelare più aggiornata esiste, ma non è conoscibile per ogni produttore. Il parametro dovrà allora essere differenziato, sussistendo due tipi di agenti modello: quello riferito a chi ha inventato la nuova tecnologia produttiva coperta da segreto e coloro che ciò non hanno fatto. Il primo produttore sarà chiamato ad adeguarsi ai nuovi standard produttivi, sia sul piano del

---

<sup>53</sup> Tesi sostenuta in dottrina da F. Mantovani, *Diritto penale, cit.*, e anche da M. Grotto, *Principio di colpevolezza, cit.* e alla quale sembra coincidere anche l'opinione di A. R. Di Landro, *Dalle linee guida e dai protocolli all'individualizzazione della colpa penale nel settore sanitario. Misura oggettiva e soggettiva della malpractice*, Torino, Giappichelli, 2012, pp. 229-312.

<sup>54</sup> Così M. Grotto, *Principio di colpevolezza, cit.*, p. 376; il quale mette in evidenza le diverse sorti cui si giungerebbe sul piano risarcitorio, quindi civilistico, laddove si aderisca alla tesi opposta, che accerta la dimensione oggettiva della colpa col parametro dell'*homo eiusdem condicionis et professionis*: il criterio dell'*homo eiusdem* escluderà la colpa già in sede di fatto tipico perché l'imprenditore non conosce ancora la regola coperta da segreto e quindi viene meno la tipicità, con la conseguenza che nessun risarcimento potrà essere accordato. Come si è già detto, l'A. aderisce alla diversa tesi che accerta la violazione della regola di diligenza con parametro della miglior scienza ed esperienza, per la quale la tipicità è integrata perché la regola cautelare è già disponibile, nondimeno il rimprovero per colpa verrà meno in sede di colpevolezza, in ragione dell'impossibilità di conoscere la predetta regola in quanto coperta da segreto. Tuttavia, la vittima avrà diritto al risarcimento, sussistendo la violazione della regola cautelare.

<sup>55</sup> (*supra*, sez. I di questo capitolo). È appunto l'impostazione della dottrina prevalente, G. Marinucci, *Innovazioni tecnologiche e scoperte scientifiche, costi e tempi di adeguamento delle regole di diligenza*, in *Riv. it. dir. proc. pen.*, 2005, pp. 40 ss.; Fiandaca, E. Musco, *Diritto penale, cit.*, p. 586; D. Pulitanò, *Diritto penale, cit.*, p. 287; D. Castronuovo, *L'evoluzione teorica, cit.*, p. 1630.

‘dovere’ impersonale, che su quello del ‘potere’ personale di agire. Il secondo, invece, non può accedere alla regola cautelare e tale impossibilità non mi pare soggettiva, ma oggettiva: è il segreto industriale che oggettivizza questa condizione e la rende uguale per tutti coloro che non sono titolari dei diritti di brevetto. La condotta di costoro non integra un fatto colposo, in quanto consiste nell’inottemperanza a una regola che neppure l’agente modello della categoria di appartenenza poteva conoscere, perché coperta da segreto. Piuttosto, sul piano oggettivo, il dovere di diligenza potrà consistere nel non aver predisposto *uno standard* produttivo più sicuro attraverso investimenti nel settore della ricerca, e non nel non essersi adeguati a *quel tipo specifico* di produzione coperto da segreto. Dopodiché, sul piano soggettivo, potrà verificarsi se l’azienda produttrice era in grado, da sola, di finanziare una ricerca per il miglioramento della sicurezza dei suoi prodotti, dando spazio a valutazioni attinenti alle peculiari circostanze del caso. Così, la colpevolezza non sarà integrata ove l’azienda si appoggi a soggetti terzi che operano nel settore della ricerca i quali non l’abbiano informata sulla possibilità di investire in nuove tecniche di produzione su cui aziende che operano nello stesso settore stavano già investendo o avevano già raggiunto dei risultati poi coperti da segreto.

V’è poi da considerare altresì l’eventualità che, a fronte di una nuova regola cautelare conoscibile da chiunque poiché non coperta da segreto, l’agente concreto sia considerato non in colpa ovvero in sede commisurativa la colpa risulti attenuata laddove vi sia stato un lasso di tempo breve affinché la persona fisica si adeguasse alla regola di condotta ovvero l’ente adeguasse il modello organizzativo<sup>56</sup>.

b.2.1) L’accertamento della misura soggettiva rileverà anche in presenza di regole cautelari scritte.

La dottrina in proposito ha notato che non è sempre vero che la possibilità soggettiva di prevedere il verificarsi dell’evento sia desumibile dall’esistenza stessa della norma<sup>57</sup>. Il caso preso in esame è quello di una regola cautelare positivizzata prevista per certi fini, tra i quali non rientri però l’impedimento dell’evento oggetto dell’incriminazione. A seguito delle scoperte scientifiche successive, si apprende però che l’osservanza di quella regola, già esistente, diffusa e conosciuta, avrebbe evitato l’evento che connota la

---

<sup>56</sup> Ipotesi ripresa dal catalogo di D. Castronuovo, *La colpa “penale”. Misura soggettiva e colpa grave*, cit..

<sup>57</sup> M. Grotto, *Principio di colpevolezza*, cit., p. 380 ss. e p. 381, nota 50, il quale fa riferimento alla violazione da parte degli imprenditori della normativa tecnica in tema di polveri allora vigente imponeva già l’utilizzo di maschere respiratorie (artt. 4, 20 e 21, D.P.R. n. 303/1956; artt. 377 e 387, D.P.R. n. 547/1955; artt. 157 e 176, D.P.R. N.1124/1965) quando ancora non si conoscevano le correlazioni tra inalazioni di polveri di amianto e patologie tumorali.

fattispecie colposa contestata. In siffatte evenienze, va condivisa l'opinione secondo la quale la prevedibilità dell'evento da parte dell'agente concreto va valutata alla luce delle conoscenze che la scienza aveva acquisito al momento del fatto; pertanto, se la regola cautelare era già esistente, ma era prevista per finalità diverse poiché non ne era ancora conosciuta l'idoneità a evitare il prodursi della lesione al bene giuridico, non può irrogarsi una condanna per non aver fatto ciò che non si sapeva ancora di dover fare per evitare quell'evento, non potendo avere rilievo le conoscenze acquisite in un momento successivo a quello della condotta<sup>58</sup>.

#### **4. La posizione dell'utente**

Tocca adesso occuparsi della posizione dell'utente.

a) Le caratteristiche intellettuali e fisiche potranno escludere o graduare la colpa dell'utente. Ad esempio, in materia di utilizzo di sistemi intelligenti saranno rilevanti le conoscenze tecniche dell'utente e la sua capacità di utilizzo di strumenti di alta tecnologia.

b) Grande rilevanza avranno le situazioni eccezionali e anomale concomitanti all'utilizzo del sistema, in grado di influenzare la riconoscibilità del rischio e l'evitabilità dell'evento da parte dell'agente concreto e di ingenerare un dubbio circa l'idoneità psico-fisica dello stesso ad agire in modo conforme alla diligenza oggettiva<sup>59</sup>.

Dovrà allora essere accertato se il comportamento negligente si è verificato in presenza di circostanze anormali, al fine di comprendere se queste hanno inciso sull'azione negligente o l'abbiano financo determinata. In questi casi la colpa del soggetto agente potrebbe risultare attenuata o persino esclusa, allorché si dimostri, rispettivamente, che sarebbe stato impossibile per l'agente concreto tenere un comportamento conforme alla regola di diligenza violata, o che tali condizioni anormali lo hanno reso molto meno padrone di se stesso che in presenza di circostanze normali. Il fatto colposo si presenta come *inevitabile* e questo fa venire meno l'elemento soggettivo del reato.

Le circostanze anormali, imprevedibili hanno allora un'incidenza sulla rilevanza penale della condotta. Sono circostanze anormali quelle concomitanti all'esecuzione di

---

<sup>58</sup> M. Grotto, *Principio di colpevolezza*, cit., p. 382.

<sup>59</sup> E. Borsatti, *La soggettivazione della colpa: profili dogmatici e pratico-applicativi*, in *Ind. pen.* 2005, pp. 75-99, in particolare p. 94.

un'attività, la cui presenza mette in dubbio la possibilità per l'agente in concreto di attenersi al comportamento prescritto dalla regola cautelare<sup>60</sup>.

Come si era già accennato, le circostanze anormali non sono sempre inquadrabili tra le circostanze che escludono la colpevolezza, avendo invece un ruolo diverso a seconda dei casi e potendo incidere su diversi momenti dell'illecito. In questo modo si crea una intersezione tra la colpa e "istituti di confine"<sup>61</sup>, nella specie il caso fortuito e la forza maggiore, e la riflessione involge altresì la nozione di azione "cosciente e volontaria". È quindi doveroso dar conto delle diverse opinioni emerse in dottrina per spiegare in quale momento dell'agire criminoso le circostanze anormali acquistino rilevanza e in che veste giuridica, avvertendo sin da ora il lettore che l'impostazione che qui si accoglie è quella che riconosce a tali istituti "di confine" un'incidenza su momenti del reato diversi da quello soggettivo, e che riconosce alla nozione di "coscienza e volontà" un ruolo nella definizione dell'azione penalmente rilevante. Proprio da quest'ultimo aspetto iniziano le riflessioni proposte nei prossimi paragrafi.

## **5. Le circostanze anormali tra la colpevolezza e il concetto di azione "cosciente e volontaria"**

Secondo la teoria dell'azione, sviluppatasi dapprima in seno alla dottrina tedesca, le circostanze straordinarie inciderebbero sulla condotta in modo da escluderne il carattere di azione penalmente rilevante, con la conseguenza di ritenere la successione degli accadimenti come 'non-azioni' per le quali non si pone il problema del contrasto con la regola di diligenza<sup>62</sup>. Il requisito della *evitabilità* si presenta allora come un *prius* rispetto

---

<sup>60</sup> E. Borsatti, *cit.*, p. 94.

<sup>61</sup> *Ibidem*, p. 96.

<sup>62</sup> G. Marinucci, *op. cit.*, p. 204. Per i sostenitori della teoria dell'azione, il concetto di "azione" è da intendersi come una superiore unità sistematica in grado di comprendere tutti i fenomeni penalmente rilevanti. La teoria dell'azione conobbe nel tempo diverse declinazioni, tutte volte a delineare un concetto "pregiuridico" di azione: accanto a una prima tesi naturalistica o causale, in cui l'azione è intesa come movimento del corpo o assenza di un movimento del corpo, si sviluppò una teoria sociale per effetto degli studi del Radbruch, per la quale l'azione coincide con la visione sociale degli avvenimenti umani. Nell'ambito di quest'ultima concezione, la difficoltà nel definire la "società" portò alla proliferazione di una teoria finalistico-sociale, o finalistica, e una teoria obbiettivo-sociale, o sociale. Per la prima, vi è azione quando l'agente abbia agito o sia rimasto inerte "volontariamente", con l'effetto però di anticipare elementi del dolo sul piano dell'azione; per la seconda, non esiste un'unica definizione del concetto 'sociale' di azione, sebene tutte mirino a individuare una visione sociale "oggettiva" dell'agire umano, così finendo per far coincidere l'azione penalmente rilevante con l'azione tipica.

Va notato che il Radbruch aveva affermato che la "pietra angolare" nella teoria del reato non è il concetto di azione, ma piuttosto il *Tatbestand* e la realizzazione dello stesso, come già aveva sostenuto E. Beling, *Methodik der Gesetzgebung, insbesondere der Strafgesetzgebung*, 1922, p. 62; nondimeno giungeva poi il

all'indagine sulla colpevolezza, forgiando un concetto unitario di azione, definita come causazione evitabile dell'evento<sup>63</sup>.

In Italia, il dibattito sul concetto di azione penalmente rilevante ruota attorno all'interpretazione dell'art. 42, comma 1, c.p. là dove prevede che “nessuno può essere punito per un'azione ... se non l'ha commessa con coscienza e volontà”<sup>64</sup>.

In particolare, è sul concetto di “coscienza e volontà” che si registrano opinioni diverse. Accanto a chi sostiene che la formula esprima un concetto unitario di azione, valevole per i reati dolosi e colposi, attivi e omissivi, v'è chi ritiene che la disposizione rilevi ai fini del giudizio di colpevolezza.

Così, secondo una prima impostazione, la norma in questione è considerata il perno del concetto unitario di azione penalmente rilevante, la quale ultima è ammantata della dominabilità del fatto da parte dell'uomo, il quale avrebbe potuto astenersi dal compiere l'azione negligente o inosservante ovvero, nelle ipotesi omissive, avrebbe potuto agire in modo diligente<sup>65</sup>. In questo modo andrebbe perciò interpretato il requisito della “coscienza e volontà” che l'art. 42, c. 1 riferisce all'azione. La dominabilità è allora requisito indispensabile per attribuire il fatto all'agente e traccia i confini della *suitas* della condotta, cioè l'attribuibilità di un certo contegno a una persona fisica, l'appartenenza dell'atto al soggetto per effetto del potenziale dominio sullo stesso da parte dell'agente<sup>66</sup>, da intendersi come “potenzialità” di attivazione dei poteri inibitori e di controllo<sup>67</sup>. La “coscienza e volontà” non si ridurrebbero allora alla descrizione del dolo, cioè dell'intenzione che accompagna l'azione umana; piuttosto descriverebbero la libertà dell'azione o dell'omissione, che si presentano, rispettivamente, come una “potenziale

---

Radbruch a conclusioni diverse, in quanto approdava sempre a un concetto di azione, anche se intesa nella veste sociale (azione come visione sociale degli avvenimenti umani) e non naturalistica o causale (azione come movimento del corpo o assenza di un movimento del corpo), G. Radbruch, *Zur Systematik der Verbrechenslehre*, in *Festgabe für R. v. Frank*, 1930, rist. 1969, p. 155. Sul punto, G. Marinucci, *op. cit.*, p. 49 ss., il cui pensiero si inquadra in quel filone che nega la possibilità gnoseologica di reperire un concetto superiore e unitario di azione, ritenendo, per contro, che la costruzione del reato inizi dalla tipicità. Sempre nella dottrina italiana, a sostenere l'importanza di un concetto unitario di azione è M. Romano, il quale legge tale nozione in modo costituzionalmente orientato, elaborando un concetto “penale-costituzionale di azione”, utile ad escludere qualsivoglia rilevanza al mero pensiero criminoso (*cogitationis poenam nemo patitur*), M. Romano, *art. 42, Commentario, cit.*, pp. 422-423.

<sup>63</sup> G. Forti, *op. cit.*, p. 189.

<sup>64</sup> Sull'argomento, si veda il lavoro di G. Balbi, *La volontà e il rischio penale d'azione*, Napoli, Jovene 1995.

<sup>65</sup> D. Castronuovo, voce *Colpa penale*, *Enc. dir., I Tematici, Il reato colposo*, Milano, Giuffrè 2021, p. 209.

<sup>66</sup> F. Antolisei, *Sul concetto dell'azione nel reato*, in *Riv. pen.*, 1925, I, p. 505; Id. *Manuale di diritto penale, P.te generale*, XII ed. a cura di L. Conti, Milano, Giuffrè, 1991, pp. 300-301; M. Romano, *art. 42, in Commentario, cit.*, p. 421.

<sup>67</sup> D. Castronuovo, voce *Colpa penale, cit.*, p. 209.

non-azione” e una “potenziale azione”: l’azione e l’omissione sarebbero “volontarie”, cioè libere, non necessitate. “Solo quando sia presente questa signoria della volontà (effettiva o potenziale) del soggetto sulla (o in relazione alla) situazione concreta, il fatto dell’uomo si distingue dal mondo dei semplici accadimenti naturali e diviene oggetto plausibile delle pretese valutative delle norme giuridico-penali”<sup>68</sup>. In sintesi, “la volontà, la quale presuppone la coscienza, quindi, nel nostro ordinamento giuridico segn[erebbe] il limite fra il reato e il non-reato”<sup>69</sup>.

Da questa lettura dell’art. 42 c. 1 c.p. si distaccano altri autori, tra cui il Marinucci, che rinuncia a leggere in questa norma il fondamento per la costruzione di una nozione superiore di azione penalmente rilevante. L’autore osserva che l’art. 42 c.1 c.p. si limita a descrivere il fatto doloso, l’unico a essere commesso con “coscienza e volontà”. Per questo, il Marinucci, per non condannare questa norma a un’interpretazione abrogante con riguardo ai fatti colposi, ritiene che essa si riferisca a *“tutte le circostanze anormali, diverse dal costringimento fisico, caso fortuito e forza maggiore, la cui presenza posseda la riconosciuta attitudine a precludere o perlomeno a rendere più difficile, anche all’agente-tipo, il potenziale intervento regolatore della coscienza e della volontà, così da rendere necessitata o almeno condizionata la realizzazione di una concomitante attività “errata””*<sup>70</sup>.

E in effetti, è proprio in relazione a quelle che il Marinucci definisce “circostanze anormali” che l’espressione di cui all’art. 42, co. 1 ha sollevato delicati interrogativi, in quanto ci si domanda se vi possa essere spazio per l’intervento penale in rapporto a condotte tenute in un momento in cui l’imputato non ha avuto una percezione e una volizione reali, cioè una responsabilità per i cosiddetti atti automatici, istintivi o riflessi e per le omissioni per dimenticanza<sup>71</sup>.

Così una parte della dottrina, il cui pensiero qui si condivide, riconosce all’art. 42, co. 1 un’attitudine selettiva delle azioni penalmente rilevanti, ritenendo che esso delinei le condizioni dell’agire personale, cioè della condotta ascrivibile all’essere umano, che presenti caratteristiche tali da renderla propria del soggetto agente; risultato, questo, al quale bisogna pervenire anche per effetto del principio della responsabilità penale

---

<sup>68</sup> M. Romano, *art. 42*, in *Commentario*, *cit.*, p. 421.

<sup>69</sup> F. Antolisei, *op. cit.*, p. 298.

<sup>70</sup> G. Marinucci, *Il reato*, *cit.*, p. 226.

<sup>71</sup> L. Eusebi, *art. 42*, in *Commentario breve al codice penale*, a cura di G. Forti, S. Seminara, G. Zuccalà, VI ed., Milano, Wolters Kluwer, 2017, p. 173 e *ivi* per una sintesi completa delle posizioni dottrinali.

personale<sup>72</sup>. Ne consegue che l'art. 42, co. 1 ha un rilievo anteriore rispetto al dolo, alla colpa e all'imputabilità e per questo la volontà intesa come "*suitas*" dovrebbe essere accertata prima della colpevolezza, almeno logicamente, se non cronologicamente<sup>73</sup>.

Al contrario, la questione della mancata responsabilità per condotte tenute in presenza di circostanze 'anormali' è stata affrontata da altra parte della dottrina osservando che in questi casi non viene in rilievo la qualità di azione di tali tipi di attività, quanto, piuttosto, il loro contrasto riprovevole con una regola di diligenza che impone una diversa attività, con la conseguenza che si non si tratta di un problema ontologico, bensì di imputazione colposa<sup>74</sup>, cioè di una questione di carattere normativo circa la rimproverabilità a titolo di colpa, che trova all'interno della colpevolezza la sua collocazione sistematica<sup>75</sup>. Per effetto dell'art. 42, co.1, allora, l'imputazione per colpa richiederebbe l'accertamento della violazione della regola di diligenza, più un ulteriore accertamento in negativo, finalizzato a constatare l'assenza di circostanze anormali che, ove presenti, avrebbero condizionato anche l'attività dell'agente modello. In altri termini, allorché in presenza di queste circostanze si rilevi che anche l'agente modello non si sarebbe attenuto al rispetto del dovere di diligenza e che la violazione dello stesso è stata resa psico-fisicamente necessitata da siffatte circostanze imprevedibili, allora non può che concludersi nel senso di escludere la colpevolezza dell'agente o, comunque, di attenuarla<sup>76</sup>. L'art. 42, co.1 sarebbe allora una disposizione generale che dà rilevanza a circostanze "innominate", aventi l'effetto di rendere psico-fisicamente necessaria l'azione contraria al dovere di diligenza, che non possono essere ignorate dall'ordinamento, il quale deve consentire l'attenuazione o l'esclusione del rimprovero per colpa.

---

<sup>72</sup> È questa la lettura di M. Romano, *art. 42*, in *Commentario*, *cit.*, p. 422-423.

<sup>73</sup> M. Romano, *art. 43*, in *Commentario*, *cit.*, p. 457.

<sup>74</sup> G. Marinucci, *cit.*, p. 208.

<sup>75</sup> Un'impostazione leggermente differente è quella di M. Gallo, il quale precisa che nei *reati colposi d'azione* la "coscienza e volontà" coincidono con la colpa dell'atto, che costituisce il 'primo gradino' perché si abbia colpa del *fatto*. L'autore, inoltre, distingue i comportamenti colposi di omissione dovuta a dimenticanza da quelli in cui l'azione è costituita da un atto automatico, osservando che solo per queste ultime la *suitas* della condotta coincide con la colpa dell'atto, cioè della fase iniziale della condotta, mentre ciò non varrebbe per le omissioni dovute a dimenticanza. In questi casi la "coscienza e volontà" non coinciderebbero con la colpa dell'atto, perché l'accertamento della violazione dell'obbligo atterrebbe al solo momento oggettivo del fatto (quindi non all'area della colpa) e, nella specie, farebbe ritenere provata l'esistenza di un'omissione rilevante per il diritto (data la presenza di un obbligo giuridico), ma alquanto rilevante sulla natura colposa o meno dell'omissione. M. Gallo, *voce Colpa penale*, *cit.*, p. 636.

<sup>76</sup> G. Marinucci, *cit.*, p. 228.

Secondo l'orientamento dottrinale in discorso, il carattere generale della disposizione in esame emergerebbe dal confronto con gli artt. 45 e 46 c.p., i quali codificano circostanze 'anormali' tipiche, la cui presenza rende psico-fisicamente necessitata l'attività negligente dell'agente: si tratta del caso fortuito, della forza maggiore e del costringimento fisico. In sintesi, nel nostro ordinamento per i reati colposi sarebbe richiesto l'accertamento di condizioni 'negative', che segue la verifica della violazione del dovere di diligenza; tale accertamento troverebbe base normativa negli artt. 45 e 46 e nella disposizione più generale che dà rilevanza anche a circostanze "innominate". Vi sarebbe quindi un legame tra gli artt. 42, co. 1, 45 e 46 c.p., in quanto si tratta di istituti escludenti il rimprovero per colpa.

Questa lettura in chiave soggettivistica del caso fortuito e della forza maggiore è da taluni condivisa, seppur con sfumature diverse, mentre altri interpretano gli istituti di cui all'art. 45 c.p. sotto altra luce. Anche qui in questa sede si aderisce<sup>77</sup> non colloca gli istituti del fortuito e della forza maggiore nel momento soggettivo del reato. È quindi giunto il momento di soffermarci su questi due istituti, che possono assumere rilevanza per l'indagine che stiamo conducendo, come a breve sarà illustrato.

## **6. Caso fortuito, forza maggiore e colpevolezza colposa**

Partiamo allora dall'analisi del caso fortuito, che viene fatto coincidere con quell'avvenimento imprevisto e imprevedibile che si inserisce d'improvviso nell'azione del soggetto, soverchiando ogni possibilità di resistenza e di contrasto, così da rendere fatale il compiersi dell'evento<sup>78</sup>. Quest'ultimo deriva quindi "dall'incrocio eccezionale della condotta con altri fattori causali, non importa se di natura umana, subumana o naturale"<sup>79</sup>.

Una diversa impostazione fa rientrare nel fortuito le circostanze anormali "interne" all'agente, come il malore rapido e improvviso, la stanchezza, la perdita parziale o totale

---

<sup>77</sup> Si tratta della concezione proposta da M. Romano, come si vedrà *infra*, par. 6.

<sup>78</sup> Cass. pen. sez. V, 28.04.1978, Negroni, in *Cass. pen.*, 1979, p. 31.

<sup>79</sup> A. Pecoraro Albani, voce *Caso fortuito (dir. pen.)*, in *Enc. dir.*, VI, Milano, Giuffrè, 1960, p. 391 ss., par. 2, il quale precisa che "Tali riflessioni inducono a rifiutare la cosiddetta concezione naturalistica del caso, secondo la quale esso sarebbe l'accadimento naturale, di per sé, prevedibile o meno (ad esempio, pioggia, alluvione, vento, morso di una vipera, ecc.)" (...), per cui "Per queste premesse, è chiaro che, ad esempio, il vento, la pioggia, il terremoto, ecc. potranno assumersi nel concetto di «caso» se non quando determinano il realizzarsi di un evento in concorso con la condotta dell'uomo. È l'incrocio tra la condotta e l'elemento naturale, che sfocia imprevedibilmente in un certo evento, a doversi qualificare caso fortuito: non la sola condotta, né il solo elemento naturale".

della coscienza<sup>80</sup>. È stato però osservato che questa opinione, seppur autorevole, restringe eccessivamente l'operatività del fortuito, dando in sostanza rilievo alle sole situazioni interne all'agente e escludendo invece fattori estranei allo stesso, che però potrebbero combinarsi in modo tale da comportare il verificarsi dell'evento<sup>81</sup>. La nozione più ampia di fortuito è da preferire, con la conseguenza che il fortuito va riferito a quelle situazioni in cui fattori interni o esterni all'agente danno luogo a un pericolo di evento che non è percepibile e quindi non è tale da poter essere controllato ed evitato con una diversa esecuzione della condotta<sup>82</sup>.

Se così è, si comprende allora come al fortuito sia ritagliato un ruolo nella delimitazione del dovere di diligenza richiesto all'agente concreto, per questo, tradizionalmente, si ritiene che il fortuito recida il legame tra il reo e l'evento in punto di colpevolezza, non potendosi in alcun modo, nemmeno a titolo di colpa, farsi risalire l'offesa all'attività psichica dell'agente<sup>83</sup>; questa impostazione è accolta anche dalla giurisprudenza, la quale, però, non manca di precisare che il fortuito è escluso laddove anche "una pur minima colpa possa essere attribuita all'agente in relazione all'evento dannoso realizzatosi"<sup>84</sup>.

Ma va dato atto dell'orientamento che sposa la concezione oggettiva del caso fortuito, ritenendo che esso rilevi sul piano del fatto tipico, escludendo la causalità, attesa l'incompatibilità tra causa e caso<sup>85</sup>.

---

<sup>80</sup> G. Marinucci, *op. cit.*, p. 224.

<sup>81</sup> V. Militello, *Rischio e responsabilità penale*, cit., p. 178, nota 80.

<sup>82</sup> V. Militello, *cit.*, p. 177.

<sup>83</sup> Si tratta dell'impostazione 'soggettivista', tra i cui sostenitori figurano F. Antolisei, *op. cit.*, pp. 336 ss.; G. Marinucci, *Il reato*, cit., p. 223.

<sup>84</sup> Cass. pen. sez. IV, 17/01/2019, n. 29922, che confermava la responsabilità ex art. 450 c.p. per il proprietario di un bovino che non aveva impedito l'invasione da parte dello stesso animale della sede ferroviaria ed il suo successivo investimento da parte di un treno, in *Diritto & Giustizia 2019*, 9 luglio (con nota di A. Ievolella); Cass. pen. sez. IV, 20/02/2015, n. 11825 che escludeva la sussistenza del caso fortuito nella condotta dell'imputata che aveva causato un incidente mortale, atteso che la donna aveva mantenuto una velocità di marcia non adeguata alle condizioni della strada, nel frangente a ridotta visibilità per la presenza di una nuvola di fumo, condizione che avrebbe imposto una ridottissima velocità, in *Diritto & Giustizia 2015*, 23 marzo.

<sup>85</sup> A. Pecoraro Albani, *op. cit.*, p. 391; F. Mantovani, *Diritto penale*, cit., p. 150 ss.. L'impostazione è invece criticata da F. Antolisei, *cit.*, p. 337, notando che l'espressione "commettere" di cui all'art. 45 c.p. presupporrebbe l'esistenza del nesso causale tra la condotta dell'uomo e il risultato. A tale critica si ribatte che sempre il dato positivo, facendo seguire la preposizione "per" al verbo "commettere", starebbe a significare "a causa di", sicché il fortuito, ma anche la forza maggiore, sarebbero *concause* dell'evento e l'art. 45 c.p. si porrebbe a integrazione dell'art. 41 c.p., dando il fortuito rilevanza alle concause preesistenti e concomitanti, mentre la forza maggiore si identificherebbe con le forze naturali sopravvenute, F. Mantovani, *ibidem*, p. 150.

In questo confronto tra opinioni diverse, è emersa una tesi mediana, che riconosce una funzione “polivalente” al caso fortuito, posto che esso interviene a volte sulla colpevolezza, escludendola, altre volte sul nesso di causalità<sup>86</sup>. Sarà allora causa escludente l’elemento soggettivo quando esso renda impossibile l’osservanza del dovere di diligenza, mentre si atteggerà a fattore di esclusione del nesso tra condotta e evento quando innescherà un rischio estrinseco rispetto alla condotta, diverso dal rischio che tipicamente sia a questa riconducibile<sup>87</sup>.

Infine, c’è chi ritiene che il fortuito, intervenendo su una condotta che è riferibile all’agente, quindi “sua”, comporti l’insussistenza del fatto tipico, cosicché si collocherebbe in un momento anteriore all’accertamento della colpevolezza<sup>88</sup>. Si distingue, allora, tra reato senza evento naturalistico, dove il fortuito determinerebbe una situazione analoga al difetto del nesso di causalità nei reati di evento, e reato con evento naturalistico, in cui un fattore imponderabile determina la condotta dell’agente. Si tratta di un’impostazione oggettivista, che però ritaglia all’art. 45 c.p. uno spazio autonomo rispetto agli artt. 40 e 41, per cui non si limiterebbe a integrarne il contenuto<sup>89</sup>.

Aderendo a quest’ultima opinione, diventa necessario distinguere quando l’inevitabilità della condotta sia dovuta a caso fortuito, con conseguente esclusione della tipicità, e quando a inesigibilità, con esclusione della colpevolezza colposa. Probabilmente la risposta sta in ciò: la colpevolezza colposa prende in considerazione solo quelle circostanze che rendono impossibile l’osservanza della regola di diligenza al singolo agente; il caso fortuito prescinde dalle capacità del singolo agente e abbraccia una costellazione di situazioni più ampia dell’inesigibilità soggettiva, per cui può dirsi che il caso fortuito è tale per tutti e non solo per chi agisce in concreto<sup>90</sup>. Pertanto, il caso fortuito abbraccia quelle circostanze anomale che rendono psico-fisicamente necessitata l’inottemperanza al dovere di diligenza non solo per l’agente concreto, ma già per l’agente modello<sup>91</sup>. Resta però fermo che i margini del caso fortuito si dissolvono quando la

---

<sup>86</sup> In questo senso, G. Fiandaca, E. Musco, *Diritto penale, cit.*, p. 235.

<sup>87</sup> È il caso della morte per l’incendio scoppiato in ospedale (caso fortuito) del soggetto ivi trasportato a seguito di un incidente stradale.

<sup>88</sup> M. Romano, in *Commentario, cit.*, sub art. 45, p. 486-87.

<sup>89</sup> *Ibidem*, p. 486.

<sup>90</sup> Si sviluppano qui le osservazioni critiche di M. Romano verso la tesi che considera il fortuito come “mancanza di dolo e colpa”, là dove l’A. osserva che questa tesi “risolvendo il fortuito nel (solo) difetto dell’elemento soggettivo, non spiega per quale ragione ciò che non è prevedibile dal singolo agente dovrebbe essere considerato fortuito dall’ordinamento (...) un fatto non doloso né colposo, ovvero non colpevole, non si intende certo normalmente come fortuito”, *cit.*, p. 485, art. 45/6.

<sup>91</sup> E. Borsatti, *cit.*, p. 96.

trasgressione del dato cautelare era prevedibile nel momento anteriore o concomitante allo svolgimento dell'attività, ovvero quando l'evento è riconducibile a una condotta soggettivamente riprovevole<sup>92</sup>.

Passando a questo punto all'esame della forza maggiore, si ritiene che essa identifichi fattori naturali (non umani, altrimenti verrebbe in gioco il costringimento fisico), o comunque circostanze "esterne" all'agente<sup>93</sup>. In presenza di forza maggiore il verificarsi del fatto è ineluttabile, perché esso è connesso *irresistibilmente* alla condotta dell'agente, mentre in presenza di caso fortuito il fatto penalmente interessante è connesso *imprevedibilmente* al contegno dell'agente<sup>94</sup>.

La forza maggiore è solitamente intesa dalla giurisprudenza come "un fatto imponderabile, imprevisto e imprevedibile" che "esula del tutto dalla condotta dell'agente, sì da rendere ineluttabile il verificarsi dell'evento", in quanto non è ricollegabile a un'azione o omissione cosciente e volontaria dell'agente<sup>95</sup>, e per questo recide "il legame psicologico tra azione ed evento, a seguito di un accertamento rigoroso"<sup>96</sup>. Su questa scia, accanto alla giurisprudenza, si colloca l'orientamento dottrinale che considera l'art. 45 c.p. una scusante al pari dell'art. 42, co.1 c.p. Secondo l'orientamento di pensiero in discorso, dall'analisi degli artt. 42, co.1, 45 e 46 emergerebbe come nel nostro ordinamento vi siano previsioni legislative che rendono psico-fisicamente necessarie certe azioni contrarie a diligenza anche per l'agente modello, che evidenziano l'insufficienza dell'accertamento della violazione della regola precauzionale e rendono necessaria l'ulteriore indagine diretta ad appurare l'assenza di circostanze "anormali" e "concomitanti"<sup>97</sup>.

---

<sup>92</sup> *Ibidem*, p. 96.

<sup>93</sup> Come già il citato Marinucci, anche M. Romano distingue tra fattori "esterni" e "interni" all'agente, riconducendo solo i primi alla forza maggiore, in *Commentario, cit., sub art. 45*, p. 487; sulla peculiare qualificazione del "malore alla guida" come circostanza non inquadrabile nella forza maggiore, né nel fortuito, ma piuttosto come circostanza che priva l'azione dei caratteri di cui all'art. 42, co. 1, *Ibidem*, p.488-89.

<sup>94</sup> A. Pecoraro Albani, *cit.*, par. 13, secondo il quale "Vari sono i criteri adottati per la differenziazione [tra caso fortuito e forza maggiore], spesso ricevuti dal diritto civile. Mentre da alcuni si dichiara che il fortuito è integrato da fatti naturali e la forza maggiore da fatti umani, da altri si sostiene l'opposto. Il criterio più convincente è quello che si riferisce alla imprevedibilità quale caratteristica del fortuito e alla irresistibilità quale nota della forza maggiore (*vis cui resisti non potest*)".

<sup>95</sup> Cass. pen. sez. III, 04/12/2007, n. 4529.

<sup>96</sup> Cass. pen. sez. III, 03/02/2021, n. 16850.

<sup>97</sup> Naturalmente la stessa indagine va poi compiuta anche per le fattispecie omissive colpose, ove oltre ad accertare l'inerzia contraria al dovere di diligenza e l'obbligo giuridico di agire, il giudizio sulla colpa non può prescindere dalla verifica dell'assenza di circostanze anormali, concomitanti all'omissione, che rendano *ex lege* l'omissione eccezionalmente necessitata e quindi scusabile, G. Marinucci, *op. cit.*, p. 232.

È qui condivisa un'opinione diversa, tendente a evidenziare come nella forza maggiore, a differenza che nel caso fortuito, mancherebbe una condotta umana cosciente e volontaria, poiché, per effetto di una causa esterna, l'agente *agitur sed non agit*<sup>98</sup>. Ne deriva che la forza maggiore esclude il requisito richiesto dall'art. 42, co.1, vale a dire una condotta cosciente e volontaria e quindi effettivamente umana. Ora, in base alla concezione che si predilige dell'art. 42 co.1, cioè come causa che esclude un'azione penalmente rilevante ovvero come causa scusante, si giunge a ritenere che la forza maggiore escluda il fatto tipico ovvero la colpevolezza<sup>99</sup>.

La giurisprudenza su questi temi è meno divisa, tendendo a escludere la colpevolezza (e non la tipicità o il carattere di azione cosciente e volontaria, cioè la *suitas*) quando è riconosciuta una causa di caso fortuito o forza maggiore incidente sulla condotta, mentre l'art. 42, co. 1 c.p. ha trovato scarsissima applicazione, risultandone vanificata l'attitudine selettiva che una parte della dottrina italiana gli riconosce<sup>100</sup>.

E, in effetti, è difficile tracciare un confine netto tra ipotesi di forza maggiore e cause di inesigibilità, pertanto, la soluzione tendente a inquadrare le circostanze anormali come

---

<sup>98</sup> M. Romano, in *Commentario, cit., sub art. 45*, p. 484 ss.; G. Fiandaca, E. Musco, *Diritto penale, cit.*, p. 233; L. Eusebi, *art. 45*, in *Commentario, cit.*, p.200; A. Pecoraro Albani, *cit.*, p. 401.

<sup>99</sup> Si rinvia in proposito alle opinioni illustrate *supra* con riferimento all'art. 42, co. 1 c.p.. Ad esempio, per M. Romano, in presenza di forza maggiore viene già meno la condotta umana propria dell'autore, ma nonostante la differenza col fortuito (che escluderebbe invece il nesso causale), in entrambi i casi vi è un'identità normativa di effetti, cioè l'insussistenza del fatto tipico, poiché l'autore legge la "coscienza e volontà" come condizioni dell'azione penalmente rilevante e non come causa di esclusione della colpevolezza. M. Romano, *art. 45*, in *Commentario, cit.*, p. 488.

<sup>100</sup> L. Eusebi, *art. 42, cit.*, p. 173, anche se proprio valorizzando l'art. 42 c.p., la giurisprudenza ha in certe occasioni riconosciuto valore scusante al malore alla guida, rimarcando l'autonomia della "coscienza e volontà" nella sistematica del reato (Cass. pen., Sez. Un., 14.06.1980, *Felloni*, CED 146695, con nota di T. Padovani). Tuttavia, va dato atto che il malore alla guida è prevalentemente ritenuto causa di esclusione della colpevolezza e classificato come caso fortuito, in quanto circostanza "interna" all'agente (*ex multis*, Cass. pen., 30.10.1990, *Lo Nigro*, CED 186075.). Considerando quindi la giurisprudenza il fortuito e la forza maggiore come incidenti in sede di colpevolezza, la classificazione di tali circostanze alla luce dell'art. 45 ovvero dell'art. 42 c.p. non determina modificazioni sostanziali del giudizio, posto che a venire meno è comunque l'imputazione a titolo di colpa del fatto, anche se l'inquadramento entro l'una o l'altra norma acquista invece rilievo sul piano probatorio per l'orientamento giurisprudenziale che pone a carico della difesa la prova del fortuito. Su questo punto, si veda il commento di G. Lunghini, L. Paris, *art. 45*, in *Codice penale commentato*, E. Dolcini, G. Marinucci, (diretto da) E. Dolcini, G. L. Gatta, Milano, Wolters Kluwer, V ed., 2021, p. 726-727, i quali sottopongono a revisione critica questo orientamento giurisprudenziale, che pone a carico della difesa la prova del fortuito qualificandolo come esimente, causa di non punibilità o elemento impeditivo del reato. Osservano i commentatori che sul piano sostanziale non può il fortuito essere qualificato come causa di non punibilità e, inoltre, l'art. 530, c. 3 c.p.p. escluderebbe una piena prova del fortuito a carico dell'imputato, poiché laddove emergano seri dubbi sull'esistenza di esso il fatto non costituirebbe già reato e l'imputato andrebbe assolto. In giurisprudenza sposa la tesi della prova a carico della difesa, tra le altre, Cassazione penale sez. IV, 18.09.2018, n. 52122, in *Guida al diritto 2019*, 8, 81, secondo la quale l'imputato "non può limitarsi ad affermare di aver dato seguito alla regola cautelare".

cause di esclusione della colpevolezza potrebbe essere un percorso più semplice per individuarne le conseguenze giuridiche.

Le affinità tra inesigibilità e forza maggiore, ad esempio, emergono dalla lettura del “*Progetto preliminare di riforma del codice penale*”, della Commissione Grosso, ove si proponeva la positivizzazione della nozione di inesigibilità, attraverso l’introduzione di una disposizione che recitava: “*La colpa è esclusa, non ostante l’oggettiva inosservanza della regola cautelare, quando l’agente si è trovato costretto ad agire, senza sua colpa, in una situazione eccezionale di panico e di fortissimo stress emotivo, tale da rendere inesigibile l’osservanza della regola*”<sup>101</sup>.

La dottrina ha subito osservato come la definizione di inesigibilità quale “situazione eccezionale di panico o di fortissimo stress emotivo” crei un forte aggancio all’istituto della forza maggiore, posto che in entrambe le ipotesi (inesigibilità e forza maggiore) si è in presenza di situazioni eccezionali che condizionano, annientandola, la *voluntas agendi* del soggetto<sup>102</sup>. Accogliendo l’orientamento che dà valore a un concetto unitario e superiore di azione e all’idea che la forza maggiore escluda proprio il carattere cosciente e volontario dell’azione, si rende necessario capire quando le “circostanze anormali” integrino un’ipotesi di forza maggiore e quando invece esse incidano sull’inesigibilità, escludendo la colpevolezza colposa. Per questo motivo la stessa dottrina si è sforzata di tracciare un distinguo e, a tal fine, ha evidenziato che in presenza di una causa di forza maggiore il soggetto incontra un ostacolo fisico e materiale, per cui egli “*non può (materialmente) scegliere di agire diversamente*”, mentre ai casi di inesigibilità è estranea la presenza di un impedimento fisico, perché il soggetto “*può (materialmente) comportarsi secondo diritto, ma ciò che manca è la capacità (psicologica) di determinarsi in conformità ad esso*”<sup>103</sup>. La diversa incidenza dogmatica delle categorie della forza maggiore e dell’inesigibilità, l’una incidente alla “capacità materiale”, l’altra alla “capacità psicologica” di agire in modo conforme alla regola cautelare non può che far propendere per la distinzione tra forza maggiore e inesigibilità, che sebbene conducano

---

<sup>101</sup> *Relazione al “Progetto preliminare di riforma del codice penale”. Parte generale (12 sett. 2000)*, in *Riv. it. dir. proc. pen.*, 2001, p. 604. La proposta non fu accolta poiché si prestava, si legge nella relazione, a introdurre profili di incertezza e discrezionalità incompatibili col principio di legalità. Sul punto è critico F. Angioni, *Note sull’imputazione dell’evento colposo con particolare riferimento all’attività medica*, in *Studi in onore di Giorgio Marinucci*, Milano, Giuffrè, 2006, p. 1279 ss., nella specie, p. 1298-99, il quale sostiene che la norma avrebbe invece rappresentato una rispettosa applicazione del principio di legalità e determinatezza.

<sup>102</sup> Così, M. Grotto, *Principio di colpevolezza*, cit., p. 353.

<sup>103</sup> *Ibidem*, p. 353.

da ultimo entrambe alla non punibilità del soggetto, lo fanno incidendo su elementi giuridici diversi che, nell'un caso escludono un'azione penalmente rilevante, in quando il soggetto *non agit, sed agitur*, nell'altro lo mettono in condizione di potersi attenere al paradigma cautelare solo a seguito di uno sforzo che è psicologicamente inesigibile.

## 7. L'inquadramento di alcuni fattori problematici: a) l'autoapprendimento

Dopo aver individuato i modi in cui certe situazioni possono incidere sul fatto di reato, escludendone il fatto tipico (caso fortuito), la colpevolezza colposa (inesigibilità), ovvero ostacolando addirittura la qualificazione di una certa azione o omissione come "cosciente e volontaria" (forza maggiore), soffermiamoci su alcuni fattori che indubbiamente giocano un ruolo nella dinamica criminosa in cui sia coinvolto l'uso di un sistema di IA, per tentarne una classificazione entro uno dei summenzionati istituti. Le situazioni su cui ci si soffermerà sono: a) l'autoapprendimento; b) la manomissione dell'algorithm; c) la cattiva o parziale informazione dell'utente; d) il malfunzionamento improvviso e incontrollabile del sistema.

Iniziamo con l'autoapprendimento.

Si tratta di una caratteristica propria dei sistemi di *machine learning* che, in sé, non può essere qualificata come improvvisa, imprevedibile, imponderabile, perché è una caratteristica essenziale del prodotto: una "prevedibile imprevedibilità"<sup>104</sup>. Imprevedibile, allora, può al più essere l'esito dell'autoapprendimento, ma non esso stesso. Non si tratterebbe di caso fortuito, né di forza maggiore perché sul piano concettuale l'apprendimento di un sistema programmato per fare ciò, ovverosia per imparare dall'esperienza, non si presenta come un fattore che non poteva in alcun modo essere immaginato dall'utente. L'utente, al contrario, è consapevole di adoperare uno strumento che può errare e che per questo deve essere sottoposto al suo controllo. In sostanza, mancherebbe nell'autoapprendimento il carattere di eccezionalità che caratterizza tanto le situazioni di caso fortuito che di forza maggiore.

Nondimeno, l'esito imprevedibile dell'autoapprendimento che determina un rischio e un evento penalmente rilevante può ben assumere rilievo ai fini del giudizio sull'esigibilità del comportamento conforme alla diligenza doverosa.

---

<sup>104</sup> Così definita da C. Piergallini, *Intelligenza artificiale, cit.*, p. 1759.

Superato positivamente l'accertamento sull'allineamento da parte dell'agente modello al dovere di diligenza pur in presenza di un apprendimento autonomo inconsueto, dovrà verificarsi se il singolo agente che aveva il controllo sul sistema in qualità di utente poteva fare altrettanto. Ciò che va valorizzato è il controllo che l'utente ha sul dispositivo intelligente: in presenza di un monitoraggio durante l'uso del sistema, anche un esito imprevedibile del suo funzionamento dovrebbe poter essere bloccato non appena si verifici. Tuttavia, bisogna, in sede di colpevolezza, tener conto dei limiti cognitivi e fisici dell'utente. Sul punto, va precisato che alcune ricerche nel campo delle interazioni uomo-macchina (*human-computer interactions "HCI"*)<sup>105</sup>, che prendono in considerazione i limiti psico-fisici degli utenti, ridimensionano di molto le capacità dell'utente-medio (aspetto che quindi potrebbe rilevare già in sede di dimensione oggettiva della colpa). In particolare, si è soliti fare riferimento al c.d. "*handover problem*"<sup>106</sup> nei veicoli a guida assistita di livello 3<sup>107</sup>, in cui è richiesto all'utente di riprendere il controllo della vettura laddove il sistema automatico riscontri delle criticità nel corso del suo funzionamento. Ebbene, tali ricerche dimostrano che mediamente il livello di attenzione con cui gli utenti monitorano la guida del sistema non è costante a causa del mancato coinvolgimento diretto degli stessi nell'atto della guida e ciò incide negativamente sulla prontezza di riflessi<sup>108</sup>. Per questo v'è anche chi sostiene che sia meglio promuovere la guida totalmente automatica, senza passare dalla guida semi-autonoma di cui al livello 3<sup>109</sup>.

Possiamo domandarci se di tali difficoltà nel livello di monitoraggio, scientificamente dimostrate, si possa tener conto ai fini dell'accertamento della colpevolezza colposa. La risposta mi pare debba essere affermativa, soprattutto nel contesto della responsabilità penale: se anche la ricerca in ambito cognitivo-psicologico

---

<sup>105</sup> A. D. Selbst, *Negligence and AI's Human Users*, in *100 Boston University Law Review*, 2020, p. 1346.

<sup>106</sup> American Association For Justice, *Driven To Safety: Robot Cars And The Future of Liability* 14 (2017), p. 14, <https://perma.cc/UNR7-56N3>, in cui si legge: "Research shows that humans are not well adapted to re-engaging with complex tasks, like driving a vehicle in an emergency situation, once their attention has been allowed to wander. A 2015 study by the National Highway Traffic Safety Administration (NHTSA) found that it took test subjects an average of 17 seconds to respond to a request to regain control of their vehicle. That's enough time for a car traveling at 60 miles per hour to travel a quarter of a mile". Il riferimento è fatto allo studio *Human Factors Evaluation of Level 2 and Level 3 Automated Driving Concepts*, National Highway Traffic Safety Administration (NHTSA), August 2015, [http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2015/812182\\_HumanFactorsEval-L2L3-AutomDrivingConcepts.pdf](http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2015/812182_HumanFactorsEval-L2L3-AutomDrivingConcepts.pdf).

<sup>107</sup> Sulla tassonomia dei veicoli a guida delegata e autonoma, si veda *infra*, cap. IV.

<sup>108</sup> V. nota precedente.

<sup>109</sup> A. D. Selbst, *cit.*, p. 1349 "That technology companies are avoiding Level 3 cars suggests that the handoff problem is so troubling because it may become the dominant cause of accidents for Level 3 cars".

dimostra che l'uomo medio non è in grado di avere sempre un livello di attenzione "straordinario", di questo il giudice dovrà tener conto nel valutare le circostanze del caso concreto<sup>110</sup> attraverso il parametro dell'esigibilità.

### 7.1 (segue) b) Manomissione e manipolazione dell'algoritmo

In un mondo sempre più connesso, i rischi cui si espone l'utilizzo dei sistemi di IA sono legati anche alla connessione alla rete, attraverso la quale soggetti terzi possono tentare di manomettere i sistemi compromettendoli con attacchi *hacker*. Ma oltre che essere manomesso, l'algoritmo può anche essere manipolato "traendolo in inganno" (*tricked*)<sup>111</sup>; la differenza è che la manipolazione avviene dall'esterno, cioè intervenendo non sul sistema, ma sui dati di *input*, sfruttando le debolezze dell'algoritmo nel riconoscimento dell'input per indurlo in errore (*physical perturbation*)<sup>112</sup>. È questo ciò che accade applicando degli adesivi a un segnale di 'stop', che non viene più riconosciuto come tale dal sistema di guida artificiale, ma considerato come un segnale di limite di velocità di 45 km/h, o ancora quanto si verifica aggiungendo un singolo adesivo a un microonde che viene per ciò solo percepito e rappresentato dal sistema di IA come un telefono<sup>113</sup>.

In questi casi si assiste a fenomeni di *adversarial machine learning*<sup>114</sup> per i quali potrà essere raggiunta la prova della responsabilità a titolo di dolo di chi ha manipolato il sistema internamente (attacco *hacker*) o dall'esterno (*physical perturbation*), mentre tale condotta potrà essere presa in considerazione per escludere o graduare la colpa dell'utente, il quale si trova a fronteggiare un rischio "imprevedibile". Più precisamente, l'inesigibilità del comportamento ossequioso della regola cautelare è motivata dalla mancata percezione da parte dell'utente degli elementi del fatto storico capaci di attivare

---

<sup>110</sup> In questo senso, sebbene nel contesto della responsabilità civile, A. D. Selbst, *cit.*, p. 1350: "because the average person cannot react in time or stay perpetually alert, failing to do so is reasonable".

<sup>111</sup> R. Calo et al., *Is Tricking a Robot Hacking?*, *University of Washington School of Law Research Paper No. 2018-05*.

<sup>112</sup> K. Eykholt et al., *Robust Physical-World Attacks on Deep Learning Visual Classification*, 2018, *arXiv.it*; sull'argomento anche A.D. Selbst, *cit.*, pp. 1350 ss.; alcuni esperimenti di manipolazione sul software di autopilot sono riportati nell'articolo di A. Bogle, *Hackers tricked a Tesla, and it's a sign of things to come in the race to fool artificial intelligence*, *ABC Science*, 14 aprile 2019, *abc.net.au*.

<sup>113</sup> K. Eykholt et al., *cit.*

<sup>114</sup> R. Calo et al., *cit.*

il percorso preventivo<sup>115</sup>, per cui la responsabilità dovrà essere provata con un'indagine sulla "visibilità" e sulla "intensità" di "circostanze-monito" (che lascino cioè supporre che il sistema sia stato manomesso o sia manipolato) che solo se effettivamente percepite completeranno l'accertamento dell'illecito colposo sul piano soggettivo<sup>116</sup>, a patto che la figura "ideale" sulla quale si valutano le capacità dell'agente concreto sia elaborata con indicazioni tratte dalla vita reale, con la consapevolezza che il sapere tecnico come quello richiesto per riconoscere un attacco *hacker* o una *physical perturbation* non sarà un sapere diffuso, essendo quello informatico un contesto in cui saranno fisiologiche sacche di incompetenza<sup>117</sup>.

## 7.2 (Segue) c) Cattiva informazione o informazione incompleta dell'utente

Il produttore è tenuto a fornire le informazioni sulle condizioni di utilizzo e sulla pericolosità del prodotto. Laddove tali informazioni siano incomplete o parziali per cause imputabili al produttore e da ciò discenda un fatto penalmente rilevante, potrà porsi un problema di accertamento della colpa dell'utente.

Abbiamo già detto che sull'utente grava il dovere di informarsi e, ragionevolmente, queste informazioni saranno messe a disposizione dal produttore. Tuttavia, per avere una conoscenza più completa del sistema di IA, l'utente potrà ricavare le informazioni anche *aliunde*. Si pensi al caso dei medici che operano con l'ausilio di robot. Il medico che si accinge ad utilizzare il macchinario sarà tenuto a conoscere le istruzioni impartite dal produttore, a effettuare un periodo di prova del macchinario, quindi un addestramento. Nondimeno, il medico potrebbe anche confrontarsi con altri colleghi che già utilizzano il robot, per avere informazioni più approfondite sulla casistica medica, oltreché effettuare degli approfondimenti consultando articoli scientifici che affrontino le complicità emerse durante l'impiego del robot.

Ma ecco il punto problematico: cosa accade se le informazioni fornite dall'azienda sono insufficienti o incomplete? Il medico viola il dovere di informarsi che compone lo standard oggettivo di diligenza, perché, come detto sopra, avrebbe potuto confrontarsi

---

<sup>115</sup> M. Grotto, *voce Scusanti, cit.*, p. 1164, classifica il mancato riconoscimento della necessità di adottare cautele comportamentali tra le scusanti "non codificate" del reato colposo.

<sup>116</sup> M. Grotto, *op. ult. cit.*, p. 1165.

<sup>117</sup> *Ibidem*, p. 1168, con riguardo al sapere tecnico che appartiene a una collettività di specialisti e non è nella disponibilità del singolo.

con altri colleghi, o comunque documentarsi tramite la lettura di articoli e report scientifici. Ma un simile livello di approfondimento da parte del medico è sempre esigibile? In particolare, se l'azienda produttrice ha fornito informazioni rassicuranti, se ha stimolato la struttura ospedaliera a incrementare l'utilizzo del robot e se ha fornito un addestramento incompleto al medico-utente, potranno probabilmente aprirsi i margini per escludere la colpevolezza colposa del medico, valorizzando il principio di affidamento in sede di dimensione soggettiva della colpa. Il medico avrebbe dovuto e potuto informarsi consultando altre fonti sui possibili rischi che derivavano dall'utilizzo del robot, quindi ha violato il dovere di diligenza. Tuttavia, se le circostanze del caso lasciano desumere che l'atteggiamento informativo dell'azienda produttrice era molto rassicurante e che l'addestramento dell'utente è stato effettuato dall'azienda mettendo a disposizione dello stesso un numero di sedute inferiori a quelle normalmente richieste, allora potranno esservi i margini per sostenere che non era esigibile da parte del medico un comportamento diverso da quello in concreto tenuto, mentre colposa sarà la condotta del produttore.

Simili questioni sono sorte nel caso *Taylor v. Intuitive Surgical, Inc.*, 187 Wash. 2d 743 (2017), giudicato dalla Supreme Court of Washington. Il caso riguardava l'utilizzo del robot da Vinci dell'azienda Intuitive Surgical per un'operazione di prostatectomia effettuata da un medico (Dr. Scott Bildsten) a un paziente (Fred Taylor), morto quattro anni dopo l'operazione, in cui era stato dimostrato il legame causale tra la morte e l'intervento. Durante l'intervento, effettuato con l'ausilio del macchinario, il medico si accorgeva che il paziente aveva delle lacerazioni interne, così decideva di proseguire l'operazione senza l'ausilio del robot. Veniva avviato un procedimento ai fini risarcitori (quindi non penale) a carico del produttore, del medico e dell'azienda ospedaliera. Un accordo risarcitorio veniva raggiunto con il medico e l'azienda ospedaliera, nei confronti dei quali il processo si estingueva, mentre proseguiva a carico dell'azienda produttrice. Con la pronuncia *Taylor v. Intuitive Surgical, Inc.*, 187 Wash. 2d 743 (2017)<sup>118</sup>, la Corte ha stabilito che grava sul produttore il dovere di informare anche l'azienda ospedaliera

---

<sup>118</sup> Per una sintesi del caso: <https://www.hortyspringer.com/documents/taylor-v-intuitive-surgical-inc-feb-2017-summary/>; la pronuncia della Corte è disponibile al link: <https://www.courts.wa.gov/opinions/pdf/922101.pdf>; per ulteriori commenti: <https://www.drugwatch.com/davinci-surgery/lawsuits/>; Andrew Villanueva, *The Legal Battle with the Future of Autonomous Surgical Robotics*, 17 *IND. HEALTH L. REV.* 367 (2020); F. Patrick Hubbard, "Sophisticated Robots": *Balancing Liability, Regulation, and Innovation*, *University of South Carolina Scholar Commons Faculty Publications Law School* 9-2014, p. 1841.

sui rischi dell'apparecchio robotico, e non solo il singolo medico che si avvarrà del robot durante l'operazione. La Corte così conclude: *“we conclude that the trial court erred by failing to instruct the jury there was a duty to warn Harrison Medical Center of the da Vinci system's risks”*.

Il caso è quindi diverso dalle riflessioni che qui si sono proposte, le quali nondimeno traggono spunto da quanto il giornale New York Times riporta sulla vicenda<sup>119</sup>, in cui si riferisce che da una corrispondenza interna all'azienda Intuitive emergeva che i rappresentanti che vendevano i robot da Vinci alle aziende ospedaliere erano invitati a presentarsi come esperti e fare pressioni affinché il robot venisse usato con standard inferiori alla norma, al fine di incrementare l'utilizzo del robot e di favorirne la vendita<sup>120</sup>.

Atteggiamenti simili da parte dell'azienda produttrice non possono non essere presi in considerazione nella valutazione della responsabilità dell'utente e, probabilmente, la sede in cui più correttamente va tenuto conto di essi è la colpevolezza colposa, trattandosi di aspetti che incidono sull'esigibilità, che possono rilevare quantomeno per graduare la colpa, influenzando sulla riconoscibilità del rischio derivante dall'impiego del macchinario e sull'effettiva possibilità di eseguire correttamente l'intervento medico.

### **7.3 (segue) d) Il malfunzionamento improvviso e incontrollabile**

*Per le strade del tredicesimo arrondissement di Parigi, un'auto dotata del sistema di guida assistita procede a elevata velocità e provoca la morte di una persona e il ferimento di altre diciannove. L'avvocato del conducente riferirà che il suo assistito aveva tentato di arrestare inutilmente il veicolo, il quale non rispondeva ai comandi, essendo il pedale del freno bloccato e mancando nella vettura un freno a mano<sup>121</sup>. Un procedimento a carico del conducente viene avviato per i reati di omicidio stradale colposo e lesioni colpose.*

---

<sup>119</sup> (<https://www.nytimes.com/2013/03/26/health/salesmen-in-the-surgical-suite.html>), articolo di Roni Caryn Rabin, (March 25, 2013),

<sup>120</sup> *“But internal company e-mails suggest that Intuitive's sales representatives were intimately involved in the process, presenting themselves as experts and pressing for lower standards in order to ease the training path for busy surgeons, to increase use of the equipment and to drive sales. In an e-mail dated May 31, 2011, a Western regional sales manager for Intuitive noted that area surgeons had used robotic equipment only five times, although the company's goal was to see 36 robotic operations performed by the end of June. He urged sales staff to persuade surgeons to switch upcoming cases to robotic ones”*

<sup>121</sup> Notizia riportata da *Le Monde*, 16 dicembre 2021, [www.lemonde.fr](http://www.lemonde.fr).

Potranno darsi non rare occasioni in cui il sistema smetterà di funzionare o andrà incontro a un blocco che lo rende non più controllabile dall'utilizzatore nel momento del suo impiego. Ebbene, tali condizioni riflettono quelle circostanze "anormali" e "concomitanti" alla violazione della regola di diligenza, con riferimento alle quali bisogna allora comprendere se sono tali da escludere la rilevanza penale della condotta umana dell'utilizzatore.

In presenza del malfunzionamento incontrollabile e improvviso del sistema di IA, la disposizione legislativa che può soccorrere in aiuto della posizione dell'utilizzatore potrebbe essere l'art. 45 c.p., con riferimento alla forza maggiore, categoria che abbraccia le circostanze esterne che si presentano come irresistibili da parte dell'utilizzatore e che perciò lo costringono necessariamente ad agire: la forza è questa volta *artificiale* e non naturale, ma comunque resta una forza esterna e irresistibile. Aderendo all'impostazione maggioritaria in dottrina che ritiene che la forza maggiore escluda qualsivoglia forma di dominio sulla situazione concreta da parte dell'agente, di talché il suo agire non sarà mai cosciente e volontario a mente dell'art. 42, co. 1 c.p., si andrà verso l'assoluzione dell'utente "perché il fatto non costituisce reato"<sup>122</sup>.

Certo, affinché la forza maggiore sia invocabile, occorre che ne sussistano i requisiti. In particolare, si ha forza maggiore quando l'agente non dispone di un sufficiente margine di scelta<sup>123</sup> e per questo è costretto ad agire. Non sempre potrà ritenersi, allora, che a fronte di un malfunzionamento del sistema di IA l'utilizzatore *agitur sed non agit*, ma si dovrà verificare se l'agente disponeva di un margine di scelta tale da escludere i presupposti dell'inevitabilità dell'evento per effetto della forza maggiore.

Nel momento in cui il malfunzionamento si presenta come un accadimento a insorgenza improvvisa e imprevedibile, stante l'assenza di preesistenti difetti di manutenzione o di anomalie osservabili<sup>124</sup>, e si intreccia con la condotta umana, allora si potrà qualificare l'avvenimento come caso fortuito, con conseguente venir meno dell'imputazione oggettiva del fatto.

---

<sup>122</sup> M. Romano, *Commentario, cit., sub art. 42/17*, p. 424.

<sup>123</sup> Sul margine di scelta che annulla i presupposti della forza maggiore, G. Fiandaca, E. Musco, *cit.*, p. 234.

<sup>124</sup> Come ritiene M. Romano, *Commentario, cit., art. 45/11*, p. 487, nel caso del blocco del volante, criticando la pronuncia Cass. 27 marzo 1980 (*Giust. pen.* 1980, II), secondo la quale "In tema di reato colposo commesso nella circolazione di autoveicolo, il blocco del volante non può costituire caso fortuito poiché non rappresenta avvenimento eccezionale, atipico e, quindi, imprevedibile".

Per colorare di maggiore concretezza il presente ragionamento, torna utile la vicenda verificatasi a Parigi con cui si è dato avvio a queste riflessioni<sup>125</sup>. La perdita del controllo della vettura dotata di un sistema di guida assistita e l'incapacità di arrestarla possono rientrare tra le cause di caso fortuito idonee a escludere il fatto tipico. E in effetti, se l'incapacità di arrestare il veicolo è dovuta al blocco del pedale del freno, alla mancata risposta del sistema a fronte di una richiesta di arresto da parte del conducente e all'incapacità di procedere diversamente al termine di quella corsa a grande velocità a causa dell'assenza di un comando manuale di blocco, cioè di un freno a mano, ci si trova dinanzi a circostanze anormali. Questo avvenimento eccezionale, oltre che improvviso, deve però anche essere imprevedibile, in maniera da impedire all'agente di adeguare la propria azione alla situazione creatasi<sup>126</sup>. Non vi sarà allora caso fortuito laddove il guidatore avesse già riscontrato problemi nel sistema di guida assistita e sperimentato malfunzionamenti simili, ancorché con effetti meno lesivi, o tali da poter innescare il sospetto di un futuro funzionamento difettoso. In siffatta evenienza, il fatto tipico è integrato e si assiste all'inottemperanza della regola cautelare sulla corretta manutenzione del veicolo.

Nel caso sottoposto al nostro esame, bisognerà allora verificare che nessun errore nell'impiego dell'auto dotata di guida assistita sia stato commesso dal conducente<sup>127</sup>. Riscontrata la tipicità, dovrà essere valutata la possibilità di scelta in capo all'agente, per verificare se gli era possibile agire diversamente in quel momento di "grande terrore", come direbbe il Marinucci, causato dall'improvvisa perdita del controllo della vettura e dall'elevata velocità con cui procedeva in una zona residenziale. A ben vedere, quella folle corsa avrebbe potuto essere arrestata direzionando la vettura contro una parete, così da evitare l'investimento degli ignari passanti, come del resto è dimostrato dall'effettivo svolgersi degli eventi: l'auto arrestava la sua corsa a seguito di uno scontro con un'altra vettura di maggiore cilindrata e quindi più pesante e in grado di impedire che la prima proseguisse il suo movimento. Ma a questo punto un ulteriore elemento non va sottovalutato: una simile manovra era *esigibile* da parte di quel conducente che scegliendo

---

<sup>125</sup> La vettura coinvolta è il modello Tesla 3, di cui si legge nell'articolo già citato « La Model 3 de Tesla est le best-seller du constructeur américain et l'une des voitures électriques les plus vendues en Europe. Le véhicule n'est pas autonome, mais est équipé d'une assistance à la conduite sur autoroute, qui impose de garder ses mains sur le volant et ses yeux sur la route ».

<sup>126</sup> Cass., sez. IV, 7 maggio 1973 – 6 febbraio 1974, n. 1058, *GP* 74, II, p. 440; *CPMA* 74, p. 749.

<sup>127</sup> Le informazioni fornite dalla stampa che ha riportato la notizia riferiscono che la vettura è dotata del sistema di guida assistita attivabile in autostrada, che richiede comunque la vigilanza del conducente nel momento in cui il sistema è attivato.

di direzionare la vettura contro una parete sarebbe andato probabilmente incontro alla morte per via dell'elevata velocità con cui procedeva la vettura? E se aggiungiamo poi che il conducente non era solo all'interno dell'auto, ma in compagnia della moglie e delle figlie: è esigibile una condotta che evita di ledere i pedoni, ma che mette a rischio la vita del conducente e della sua famiglia?

Si apre spazio per l'invocazione dello stato di necessità quale scusante legalmente riconosciuta quando il pericolo è riferito allo stesso agente o a un suo prossimo congiunto<sup>128</sup>, sostenendo che la condotta alternativa avrebbe evitato la lesione dei beni della vita e dell'integrità fisica di terzi, ma avrebbe sacrificato quelli dell'agente e della sua famiglia. Ma, ancora una volta, la scusante non potrà operare in caso di pericolo volontariamente causato, da intendersi tale anche quello che derivi da condotta colposa, e non solo dolosa<sup>129</sup>. Bisognerà allora appurare in sede di indagini se il conducente aveva attivato il sistema di guida assistita in condizioni diverse da quelle consentite, cioè in un contesto urbano in cui non poteva essere attivato, essendo per contro consentito solo nei tratti autostradali. Se così fosse, non potrà di certo avere efficacia scusante la circostanza 'anormale' che lo stesso agente ha creato, non essendovi margini per invocare il caso fortuito o l'art. 42 c.1, c.p. poiché, sostituendo all'agente in concreto l'agente modello, si appurerebbe che il secondo avrebbe tenuto un comportamento differente, dal quale l'agente in concreto si è discostato, così attirandosi il rimprovero per colpa ed escludendo l'operatività di una scusante.

Un brevissimo cenno va però fatto alla posizione del produttore. Una responsabilità da prodotto intelligente difettoso andrebbe incontro alle difficoltà probatorie che sin qui sono emerse, soprattutto in tema di causalità. Nondimeno, una responsabilità può esservi per non aver installato un sistema manuale di arresto del veicolo<sup>130</sup>. Collocandoci nell'ottica di sistemi di IA antropocentrici, che richiedono comunque un controllo dell'uomo, non può di certo ritenersi conforme a tali requisiti il manufatto privo di un

---

<sup>128</sup> A favore della lettura dell'art. 54 c.p. come scriminante, con riferimento al soccorso di necessità, e come scusante per ciò che concerne il coinvolgimento della vita dello stesso agente o di un prossimo congiunto, G. Fiandaca, E. Musco, *Diritto penale, cit.*, p. 321 e p. 432. In generale, sullo stato di necessità, cfr. F. Viganò, *Stato di necessità e conflitti di doveri. Contributo alla teoria delle cause di giustificazione e delle scusanti*, Milano, Giuffrè, 2000.

<sup>129</sup> Così, Grosso, *Difesa legittima e stato di necessità*, Milano, Giuffrè, 1964, p. 99 ss.; M. Romano, *cit. sub art. 54/10* p. 572; G. Fiandaca, E. Musco, *cit.*, p. 323.

<sup>130</sup> Ma anche la mancanza di sistemi di arresto automatico di emergenza (quindi che non richiedono l'intervento umano) può costituire motivo di responsabilità del produttore. B. W., Smith, *Automated Driving and Product Liability, cit.*, p. 49.

sistema di controllo di emergenza. Questa mancanza dà essenza alla colpa per violazione del dovere di diligenza, consistente nel produrre e immettere in commercio sistemi sicuri e affidabili e sottoposti comunque a un controllo umano significativo, che sia in grado di governare il veicolo a fronte di condizioni improvvise di malfunzionamento.



## CAPITOLO IV

### DISCIPLINARE L'INTELLIGENZA ARTIFICIALE.

#### LA RIFORMA FRANCESE SULLA RESPONSABILITÀ PENALE DA USO DI AUTO A GUIDA AUTONOMA

*Sommario:* Premessa - 1. La sistematica del reato nel diritto penale francese - 1.1 Il principio di legalità e lo strumento delle *ordonnances* - 1.2 L'elemento materiale nei reati colposi - 1.2.1 La causalità diretta tra volontà del legislatore e interpretazione della giurisprudenza - 1.2.2 Il nesso di causalità indiretto - 1.2.3 Certezza del nesso di causalità - 1.3 *La faute* - 1.3.1 L'elemento morale nelle contravvenzioni - 2. Responsabilità penale e veicoli a guida autonoma. Alcune riflessioni sulla recente riforma del *code de la route* - 2.1. *Code de la route* e *code pénal* - 3. La riforma del 2021 e la disciplina della responsabilità penale in caso di utilizzo di auto a guida autonoma - 4. La responsabilità del conducente - 5. La responsabilità del produttore - 5.1 Il regime delle contravvenzioni. Responsabilità per il fatto del terzo? - 5.2 Il dibattito sorto intorno all'art. 121-3 del *code de la route* e la compatibilità con i principi della responsabilità personale e della presunzione di innocenza - 5.3 La lettura della norma sulla responsabilità del produttore alla luce della pronuncia del *Conseil constitutionnel* del 1999 - 6. Un esempio da imitare?

#### Premessa

Tra le applicazioni più diffuse dell'intelligenza artificiale in grado di sollecitare in modo immediato la riflessione giuridica vi è quella delle auto a guida autonoma, probabilmente perché si tratta di una tecnologia che si presta a essere a disposizione di molti individui nel prossimo breve periodo.

Quando si discute di guida autonoma si suole fare riferimento alla classificazione proposta dalla Society of Automotive Engineers (SAE), che individua in totale sei livelli di autonomia nella guida secondo un ordine crescente di automazione che va dal livello 0 (*no driving automation*), al livello 5 (*full driving automation*), passando per i livelli 2 (*driver assistance*), 3 (*partial driving automation*, in cui il conducente deve riprendere la guida del veicolo su richiesta del sistema) e 4 (*high driving automation*)<sup>1</sup>. Le questioni legate alla guida autonoma si pongono a partire dal livello 3, posto che i livelli 1 e 2 descrivono i sistemi di guida assistita (il sistema regola il livello di velocità e, solo in presenza di determinate circostanze, anche la direzione di marcia) e non autonoma.

La Francia è uno degli Stati europei in cui il dibattito sulla vendita e sull'utilizzo delle auto a guida autonoma è in atto ormai da tempo e ciò ha stimolato il legislatore francese ad elaborare un quadro legislativo favorevole all'industria automobilistica<sup>2</sup>,

---

<sup>1</sup>*Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, [www.sae.org](http://www.sae.org), aggiornato al 30.04.2021.

<sup>2</sup> S. Hadri, *Pleins phares sur les évolutions du cadre légal et réglementaire des véhicules autonomes*, *Petites affiches*, n. 3, 3 janv. 2019, p. 6.

quantomeno perché inizia a fare chiarezza su questioni che presto si porranno con una certa frequenza.

La sperimentazione di veicoli a guida delegata era stata autorizzata già nel 2016 con l'*ordonnance* n. 2016-1057 del 3 agosto 2016. Tuttavia, è con la legge n. 2019-486 del 22 maggio 2019 che il legislatore francese ha impresso un'accelerazione alla trasformazione dei mezzi di trasporto. La legge da ultimo citata è stata adottata nell'ambito del *plan d'action pour la croissance et la transformation des entreprises* (PACTE), piano che si prefigge un rinnovamento delle imprese del paese, e per questo detta "loi PACTE". Tra i settori coinvolti nella loi PACTE vi è anche quello dell'utilizzo sperimentale di auto a guida autonoma, attraverso la modifica dell'*ordonnance* n. 2016-1057 del 3 agosto 2016 che aveva già consentito la sperimentazione dei veicoli a guida delegata sopra citata; in particolare, la loi PACTE ne autorizza la circolazione su strada previo rilascio di un'autorizzazione<sup>3</sup>.

Questo intervento normativo è interessante ai fini del presente lavoro poiché esso ha delle ricadute anche in materia di responsabilità penale del conducente, che è esclusa allorché il sistema è attivato ed è impiegato in modo conforme alle istruzioni per l'utilizzo; la responsabilità ricade allora sul titolare dell'autorizzazione di sperimentazione, sia per le violazioni delle norme del codice della strada, che per omicidio e lesioni stradali, ferme restando le condizioni di cui all'art. 121-3 c. pén.<sup>4</sup> (norma che, come si vedrà dappresso, fissa i criteri di attribuzione della responsabilità). Il conducente è nuovamente

---

<sup>3</sup> Art. 125 loi du 22 mai 2019 : I.- L'*ordonnance* n° 2016-1057 du 3 août 2016 relative à l'expérimentation de véhicules à délégation de conduite sur les voies publiques est ainsi modifiée :

1° L'article 1er est ainsi rédigé : « Art. 1.-La circulation sur la voie publique de véhicules à délégation partielle ou totale de conduite à des fins expérimentales est autorisée. Cette circulation est subordonnée à la délivrance d'une autorisation destinée à assurer la sécurité du déroulement de l'expérimentation.

« La délivrance de l'autorisation est subordonnée à la condition que le système de délégation de conduite puisse être à tout moment neutralisé ou désactivé par le conducteur. En l'absence de conducteur à bord, le demandeur fournit les éléments de nature à attester qu'un conducteur situé à l'extérieur du véhicule, chargé de superviser ce véhicule et son environnement de conduite pendant l'expérimentation, sera prêt à tout moment à prendre le contrôle du véhicule, afin d'effectuer les manœuvres nécessaires à la mise en sécurité du véhicule, de ses occupants et des usagers de la route ». per un commento in dottrina, M. Bénéjat, *Le droit pénal de la route face aux nouveaux modes de transport*, *AJ Pén.*, sept. 2019, pp. 428-430 ; L. Teresi, *Véhicule à délégation de conduite et risque automobile : une lecture juridique*, 2020, *actu-juridique.fr*.

<sup>4</sup> Nuovo art. 2-2 ord. n. 2016-1057 du 3 août 2016: « Si la conduite du véhicule, dont le système de délégation de conduite a été activé et fonctionne dans les conditions prévues au premier alinéa de l'article 2-1, contrevient à des règles dont le non-respect constitue une contravention, le titulaire de l'autorisation est pécuniairement responsable du paiement des amendes. Si cette conduite a provoqué un accident entraînant un dommage corporel, ce titulaire est pénalement responsable des délits d'atteinte involontaire à la vie ou à l'intégrité de la personne prévus aux articles 221-6-1, 222-19-1 et 222-20-1 du code pénal lorsqu'il est établi une faute au sens de l'article 121-3 du même code dans la mise en œuvre du système de délégation de conduite ».

responsabile allorché il sistema ne abbia richiesto l'intervento e il primo sia rimasto inerte<sup>5</sup>.

L'intervento del legislatore francese apparentemente si pone in contrasto con la Convenzione di Vienna del 1968 (ratificata dalla Francia il 9 dicembre 1971), secondo la quale ogni veicolo in movimento deve avere un conducente che ne abbia costantemente il controllo<sup>6</sup>. Tale principio resta intatto anche dopo le timide modifiche<sup>7</sup>, relative ai sistemi di bordo che influiscono sulla guida del veicolo<sup>8</sup>, apportate dagli emendamenti del 26 marzo 2014, in vigore dal 23 marzo 2016 (RU 2016-1019)<sup>9</sup>. Così sembra che la necessità di un conducente alla guida, richiesta dalla Convenzione, rappresenti un ostacolo all'introduzione di tecnologie più avanzate.

Nondimeno, proprio al fine di conciliare lo sviluppo tecnologico con l'apparato normativo internazionale, la dottrina non ha mancato di evidenziare che se è vero che la Convenzione di Vienna non autorizza l'impiego di sistemi totalmente privi di un conducente<sup>10</sup>, come ad esempio i robot-taxi, per contro essa non intralocerebbe lo sviluppo di sistemi di guida assistita o autonoma che richiedono la vigilanza di una persona fisica. Per sostenere ciò si fa leva sul fatto che la Convenzione non fornisce la definizione di conducente, né tantomeno indica il luogo in cui esso debba trovarsi, cosicché si ritiene che è conducente la persona fisica che si trovi tanto all'interno del veicolo a motore, quanto all'esterno dello stesso, sempreché ne eserciti il controllo<sup>11</sup>. Tuttavia resta fermo

---

<sup>5</sup> Nuovo art. 2-1 ord. n. 2016-1057 du 3 août 2016 : «*Le premier alinéa de l'article L. 121-1 du code de la route n'est pas applicable au conducteur pendant les périodes où le système de délégation de conduite, qu'il a activé conformément à ses conditions d'utilisation, est en fonctionnement et l'informe en temps réel être en état d'observer les conditions de circulation et d'exécuter sans délai toute manœuvre en ses lieux et place.*

«*Le même premier alinéa est à nouveau applicable après sollicitation du système de conduite et à l'issue d'un délai de reprise de contrôle du véhicule précisé par l'autorisation d'expérimentation, dont le conducteur est informé. Il en va de même lorsque le conducteur a ignoré la circonstance évidente que les conditions d'utilisation du système de délégation de conduite, définies pour l'expérimentation, n'étaient pas ou plus remplies.*

<sup>6</sup> Conv. Vienna, art. 8 par. 1 e 5.

<sup>7</sup> S. Hadri, *op. cit.*

<sup>8</sup> Art. 5bis. *I sistemi di bordo che influiscono sulla guida del veicolo sono considerati conformi al paragrafo 5 del presente articolo e al primo paragrafo dell'articolo 13 se sono conformi alle disposizioni in materia di costruzione, montaggio e utilizzo previste negli strumenti giuridici internazionali riguardanti i veicoli a ruote e gli equipaggiamenti e componenti montati e/o utilizzati sugli stessi. I sistemi di bordo che influiscono sulla guida del veicolo e non conformi alle disposizioni in materia di costruzione, montaggio e utilizzo summenzionate sono considerati conformi al paragrafo 5 del presente articolo e al primo paragrafo dell'articolo 13 se possono essere neutralizzati o disattivati dal conducente.*

<sup>9</sup> Sul punto, C. Coulon, *Révision de la Convention de la Circulation routière : les voitures autonomes (pas tout à fait) sur la ligne de départ*, *Resp. civ. et assurance*, n. 6, alerte 7, 2016.

<sup>10</sup> I. Vingiano-Viricel, *Véhicule autonome: qui est responsable ?*, Paris, Lexis Nexis, 2019, p.11.

<sup>11</sup> I. Vingiano-Viricel, *op. cit.*, p. 15-16.

che conducente è solo la persona fisica, mentre non può essere tale la persona giuridica, né tantomeno un sistema di guida<sup>12</sup>.

L'attenzione rivolta dal legislatore francese ai sistemi a guida autonoma non si è arrestata alla sola fase di sperimentazione. Nel 2021 con una *ordonnance* è intervenuto a disciplinare in generale l'utilizzo dei sistemi di guida delegata, modificando il Code de la route e prevedendo un'apposita disciplina in tema di responsabilità penale.

Così, la Francia è oggi tra i primi Paesi europei ad aver adottato una legge in materia di guida autonoma, ma non l'unico.

Bisogna infatti ricordare che in Germania è in atto l'approvazione del disegno di legge che modifica il codice della strada permettendo alle auto a guida autonoma di essere utilizzate in tutto il paese, già approvato dal Bundestag (il Parlamento federale) e attualmente in discussione in seno al Bundesrat (il Senato federale), che autorizzerebbe la circolazione di auto a guida autonoma di livello 4 della classificazione SAE<sup>13</sup>, livello non ancora autorizzato in Francia<sup>14</sup>. Il livello 4 è ancora compatibile con la Convenzione di Vienna che richiede la presenza di un conducente nei termini di cui si è detto poco sopra<sup>15</sup> e ciò ne consente la sperimentazione nei Paesi firmatari.

In Italia è attualmente consentita l'omologazione di veicoli a guida assistita di livello 2<sup>16</sup>, mentre le questioni legate alla guida autonoma si pongono a partire dal livello 3, posto che i livelli 1 e 2 descrivono i sistemi di guida assistita (ad esempio il regolatore del livello di velocità) e non autonoma<sup>17</sup>.

Tornando alla legislazione francese, l'*ordonnance* di modifica del code de la route riguarda il livello 3 di guida autonoma, che provvede autonomamente al controllo del traffico in situazioni e contesti predeterminati (come ad esempio la guida in autostrada), ma richiede sempre la presenza di un conducente che vigili sull'operato del sistema e ripristini all'occorrenza il controllo manuale<sup>18</sup>. Essa si inserisce in un contesto fattuale in

---

<sup>12</sup> *Ibidem*, p. 15.

<sup>13</sup> G. Marino, *La Germania dice sì alla guida autonoma: è la prima in Europa*, [www.repubblica.it](http://www.repubblica.it), 31 maggio 2021.

<sup>14</sup> S. Hadri, *op. cit.*

<sup>15</sup> I. Vingiano-Viricel, *op. cit.*, p. 16, mentre sarebbero vietati i mezzi totalmente privi di un conducente anche all'esterno del veicolo, come nel caso dei "robot-taxi".

<sup>16</sup> Decreto 28 febbraio 2018, *su modalità attuative e strumenti operativi della sperimentazione su strada delle soluzioni di Smart Road e di guida connessa e automatica (c.d. Decreto Smart Road)*, GU serie gen. n. 90 del 18.04.2018, [www.gazzettaufficiale.it](http://www.gazzettaufficiale.it)

<sup>17</sup> *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, [www.sae.org](http://www.sae.org), aggiornato al 30.04.2021

<sup>18</sup> G. Torchiani, *Auto a guida autonoma: cosa sono e come funzionano*, 18 maggio 2021, [www.ai4business.it](http://www.ai4business.it).

cui l'uso delle auto a guida autonoma non è ancora diffuso, poiché finora tale impiego è avvenuto solo in via sperimentale dopo l'approvazione della loi PACTE. Si tratta, quindi, di un raro (o forse unico?) caso in cui il diritto anticipa la tecnologia, attraverso una disciplina elaborata sulla base di previsioni, immaginando situazioni potenzialmente offensive che potrebbero presentarsi in futuro, una volta che la tecnologia già esistente inizierà a essere impiegata.

Prima di analizzare i contenuti della recente riforma, è necessario accennare ai criteri che guidano l'accertamento della responsabilità colposa nell'ordinamento penale francese, per poi passare a un'introduzione al code de la route in generale, che analizzi altresì i rapporti con il code pénal. Di questo si occupa la prima parte di questo lavoro, mentre la seconda parte è dedicata alla riforma del 2021, soffermandosi su alcune questioni critiche che la stessa solleva.

## **1. La sistematica del reato nel diritto penale francese**

### **1.1 Il principio di legalità e lo strumento delle *ordonnances***

A partire dalla rivoluzione francese, l'ordinamento ha fatto propria la classificazione tripartita dei reati (*infractions*) che, sulla base della gravità discendente, si è soliti presentare nell'ordine seguente: crimini (*crimes*), delitti (*délits*) e contravvenzioni (*contraventions*), oggi recepita nel primo articolo del code pénal (art. 111-1)<sup>19</sup>.

In materia penale vale il principio di legalità che però fa riferimento non solo alla fonte legislativa, ma anche quella secondaria, cioè i regolamenti. In proposito l'art. 111-2 c. pén. afferma che i crimini e i delitti sono stabiliti dalla legge, mentre le contravvenzioni trovano fonte nei regolamenti<sup>20</sup>. Per questo motivo il code pénal è diviso in due parti: una di rango legislativo, i cui articoli sono preceduti dalla lettera "L", e l'altra di rango regolamentare, evincibile dalla presenza della lettera "R" davanti a ciascun articolo.

---

<sup>19</sup> Art. 111-1 c. pén. *Les infractions pénales sont classées, suivant leur gravité, en crimes, délits et contraventions.* Sul superamento della concezione tripartita in favore di quella bipartita, che accomuni i crimini e i delitti per isolare le sole contravvenzioni, in ragione delle affinità tra le prime due, si veda Y. Mayaud, *Droit pénal général*, 6 ed., Paris, Presses Universitaires de France (PUF), 2018, pp. 350 ss.

<sup>20</sup> Art. 111-2 c. pén : *La loi détermine les crimes et délits et fixe les peines applicables à leurs auteurs. Le règlement détermine les contraventions et fixe, dans les limites et selon les distinctions établies par la loi, les peines applicables aux contrevenants.*

La delimitazione del campo di applicazione della legge e dei regolamenti è stabilita dalla Costituzione della V Repubblica, in particolare, l'art. 34 cost. fr. stabilisce le materie di competenza legislativa in cui la legge fissa regole certe e i principi fondamentali; tutte le materie non menzionate sono di competenza regolamentare, come previsto dal successivo art. 37 Cost., di talché si desume che le contravvenzioni non trovano fonte nella legge, ma nei regolamenti.

Si distinguono regolamenti adottati in applicazione di una legge, che intervengono nelle materie indicate all'art. 34 Cost. e sono subordinate a quella legge, e i regolamenti autonomi, che intervengono nelle materie residuali e che si pongono allo stesso livello della legge sul piano della gerarchia delle fonti<sup>21</sup>. Da ciò discende che la fonte regolamentare ha uno spazio applicativo più esteso, in quanto residuale, tanto è vero che in Francia sono in vigore più di 120 mila regolamenti e solo 10 mila leggi<sup>22</sup>.

Questa distinzione costituzionale si rispecchia, come si avrà modo di vedere più avanti, nel *Code de la Route*, in cui vi è una parte legislativa e una di fonte regolamentare.

Nella gerarchia delle fonti, bisogna poi distinguere tra *leggi formali*, votate dal Parlamento, e *leggi materiali*, che sono dei decreti e, per questo, sono alle prime subordinate.

Vi sono poi le *ordonnances*, di cui è bene dar brevemente conto poiché è questo lo strumento normativo utilizzato per introdurre le nuove regole in tema di responsabilità del conducente e del produttore di auto con sistema di guida autonoma. Nella specie si tratta di provvedimenti adottati dal Governo in applicazione di una *loi d'habilitation* votata dal Parlamento, che fissa altresì il termine entro il quale dovrà procedersi alla ratifica dell'*ordonnance* da parte del Parlamento, come previsto dall'art. 38 Cost.<sup>23</sup>. Tale potere si esplica in materie che normalmente sarebbero di competenza legislativa. Si tratta di uno strumento per il quale sorgono delle tensioni col principio di riserva di legge, così

---

<sup>21</sup> R. Cabrillac, *Introduction générale au droit*, Paris, Dalloz, 2021, p. 114.

<sup>22</sup> J.L. Halpérin, *Introduction au droit*, Paris, Dalloz, 2019, p. 14

<sup>23</sup> Art. 38 Cost. Fr. *Le Gouvernement peut, pour l'exécution de son programme, demander au Parlement l'autorisation de prendre par ordonnances, pendant un délai limité, des mesures qui sont normalement du domaine de la loi.*

*Les ordonnances sont prises en conseil des ministres après avis du Conseil d'État. Elles entrent en vigueur dès leur publication mais deviennent caduques si le projet de loi de ratification n'est pas déposé devant le Parlement avant la date fixée par la loi d'habilitation. Elles ne peuvent être ratifiées que de manière expresse.*

*A l'expiration du délai mentionné au premier alinéa du présent article, les ordonnances ne peuvent plus être modifiées que par la loi dans les matières qui sont du domaine législatif.*

come nel nostro ordinamento lo stesso problema è posto dall'uso del decreto legislativo e del decreto legge in materia penale.

Le *ordonnances* entrano in vigore a partire dalla loro pubblicazione, ma perdono efficacia se non vengono ratificate in modo espresso dal Parlamento entro il termine stabilito dalla *loi d'habilitation*. Tuttavia, una recente decisione del *Conseil constitutionnel* è intervenuta a precisare la natura giuridica delle *ordonnances* non ratificate<sup>24</sup>. La pronuncia riguarda il caso di una *ordonnance* per la quale era già stato depositato il progetto di ratifica nei termini previsti dalla *loi d'habilitation*, ciononostante il Parlamento non aveva ratificato l'*ordonnance*. Ebbene, il *Conseil constitutionnel* ha avuto modo di precisare che spirato il termine stabilito dalla *loi d'habilitation* per la ratifica, a mente dell'art. 38 Cost. fr., le disposizioni contenute nell'*ordonnance* possono essere modificate solo dalla legge qualora interessino materie di competenza legislativa. Da ciò la pronuncia fa discendere che a far data dalla scadenza del termine per la ratifica, il contenuto dell'*ordonnance* acquista valore legislativo<sup>25</sup>. La vigenza delle *ordonnances* non ratificate per le quali sia stato presentato un progetto di ratifica non approvato è stata ribadita in una successiva pronuncia<sup>26</sup> in cui il *Conseil constitutionnel* ha altresì affermato che le *ordonnances* possono essere oggetto di una questione di legittimità costituzionale (a conferma della loro assimilazione alla legge) e tale sindacato può persino estendersi anche alla *loi d'habilitation*.

Va dato atto del fatto che è raro nella prassi che il Parlamento, in sede di ratifica, apporti delle modifiche al testo dell'*ordonnance* tramite emendamenti. La ratifica si riduce così a una mera formalità e ciò probabilmente attenua la portata della decisione del

---

<sup>24</sup> Conseil constitutionnel, 28 mai 2020 (n° 2020-843 QPC), [www.conseil-constitutionnel.fr](http://www.conseil-constitutionnel.fr) ; Verpeaux, M., *Actes administratifs - Ordonnances = dispositions législatives. A propos de la décision du Conseil constitutionnel du 28 mai 2020* », *La Semaine juridique. Administrations et collectivités territoriales*, 22 juin 2020, n° 25, p. 4 ; Levade, A., *Conseil constitutionnel et ordonnances : l'in vraisemblable revirement !*, *La Semaine juridique. Édition générale*, 29 juin 2020, n° 26, p. 1185 - 1186

<sup>25</sup> Par. 11 « si un projet de loi de ratification de l'ordonnance du 5 août 2013 a été déposé dans le délai fixé par l'article 12 de la loi du 27 décembre 2012, le Parlement ne s'est pas prononcé sur cette ratification. Toutefois, conformément au dernier alinéa de l'article 38 de la Constitution, à l'expiration du délai de l'habilitation fixé par le même article 12, c'est-à-dire à partir du 1er septembre 2013, les dispositions de cette ordonnance ne pouvaient plus être modifiées que par la loi dans les matières qui sont du domaine législatif. Dès lors, à compter de cette date, elles doivent être regardées comme des dispositions législatives ».

<sup>26</sup> Conseil constitutionnel, 3 juillet 2020 (n° 2020-851/852 QPC), [www.conseil-constitutionnel.fr](http://www.conseil-constitutionnel.fr) , par. 10 : « dès lors qu'un projet de loi de ratification a été déposé devant le Parlement avant la date fixée par la loi d'habilitation, les ordonnances demeurent en vigueur y compris si le Parlement ne s'est pas expressément prononcé sur leur ratification » ; A. Levade, *Ordonnance et loi d'habilitation, le Conseil constitutionnel étend son empire*, *La semaine juridique - édition générale - n° 30-35 - 27 juillet 2020*

*Conseil constitutionnel* tesa a equiparare le *ordonnances* non ratificate (ma per le quali sia stata depositato un progetto di ratifica) alla legge.

Fatte queste premesse in materia di fonti che ci torneranno utili nel prosieguo, occupiamoci adesso della sistematica del reato, che anche nel sistema francese si caratterizza per la presenza di un elemento materiale e uno morale.

## 1.2 L'elemento materiale nei reati colposi

L'accertamento del nesso di causalità nel diritto penale francese è strettamente connesso alla colpevolezza, e in ciò il sistema francese differisce da quello italiano.

In passato l'accertamento del nesso causale non conosceva differenze tra la responsabilità civile e quella penale: il criterio era unico e uniforme. Inoltre, la responsabilità penale conseguiva all'accertamento di un nesso di causalità sia diretto, che indiretto, senza distinzione alcuna tra reati dolosi e colposi, facendo in sostanza applicazione della teoria dell'equivalenza delle condizioni<sup>27</sup>.

L'ammissione di un nesso di causalità indiretto anche per i reati colposi ha però conosciuto un temperamento con la legge del 10 luglio 2000 (loi n. 2000-647, detta Loi Fauchon) che è intervenuta proprio in tema di causalità e di reati colposi modificando l'art. 121-3 c. pén<sup>28</sup>, allo scopo di “*mettre fin à la confusion entre la faute civile et la faute pénale d'imprudence, en exigeant une faute caractérisée pour que la responsabilité pénale d'une personne physique soit engagée*”<sup>29</sup>. Per questo la legge del 10 luglio 2000 interveniva per delimitare la responsabilità penale dei soggetti chiamati ad assumere un ruolo decisivo e organizzativo, soprattutto in contesti pubblici, senza tuttavia frustrare le esigenze di tutela emergenti dai settori della circolazione stradale e della sicurezza nei luoghi di lavoro<sup>30</sup>.

---

<sup>27</sup> F. Desportes, F. Le Guehec, *Droit pénal général*, 6ed., Paris, Economica, p. 481 ; Y. Mayaud, *Droit pénal général*, 6ed., Paris, Presses Universitaires de France, 2018, p. 407. *Ex multis*, in tema di lesioni dolose e causalità diretta, *Crim. 27 févr. 1992, Dr. Pén., 1992, comm. N. 199* ; in tema di omicidio e lesioni colpose e causalità diretta *Crim. 20 juin 1989, Dr. Pén. 1989, comm 60; Cass. crim. 18 oct. 1995, Bull. Crim. N. 314*.

<sup>28</sup> Per un commento, v. D. Commaret, *La loi Fauchon, cinq ans après, Droit pénal n° 4, Avril 2006, étude 7* ; P. Conte, *Le lapiste et la mort, Dr. Pén. n° 1, Janvier 2001, chron. 2* ; B. Cotte, D. Guihal, *La loi Fauchon, cinq ans de mise en œuvre jurisprudentielle, Dr. pén. n° 4, Avril 2006, étude 6* ; P. Conte, *Le lapiste et la mort, Dr. Pén. n° 1, Jan. 2001, chron. 2*.

<sup>29</sup> Proposition de loi sur la définition des délits non intentionnels, RAPPORT 391 (1999-2000) - commission des lois, Senat, [www.senat.fr](http://www.senat.fr).

<sup>30</sup> D. Commaret, *La responsabilité pénale des décideurs en matière de délits non intentionnels depuis la loi du 10 juillet 2000, Gaz. Pal. N. 255, 11.09.2004, p. 3*.

Tale obiettivo fu perseguito intervenendo su un duplice piano: da un lato su quello della causalità, dall'altro su quello della gravità della colpa. Il risultato è che l'art. 121-3 rappresenta oggi una norma centrale nello studio della causalità e della colpa poiché, dopo aver distinto i vari gradi della colpa, intesse una stretta relazione tra ciascuno di essi e la causalità. E infatti, in caso di colpa dovuta a imprudenza, o negligenza, che per esemplificazione espositiva la manualista definisce *faute simple*, è richiesto un nesso di causalità *directo*. Mentre, laddove il nesso di causalità è indiretto la *faute simple* non è sufficiente, ma occorre un atteggiamento colposo corrispondente, alternativamente, alla *faute délibérée*, che potremmo accostare alla nozione di colpa cosciente nel nostro ordinamento, ovvero alla *faute caractérisée*. Così, il carattere diretto o indiretto del nesso di causalità vincola l'interprete nell'accertamento del grado della colpa: se la causalità è diretta è sufficiente una colpa semplice; se la causalità è indiretta serve una colpa "aggravata"<sup>31</sup> nelle due manifestazioni di cui si è appena detto.

Se di questi tipi di colpa si tratterà nel paragrafo successivo, bisogna adesso capire cosa si intende per causalità diretta e indiretta, tenendo a mente che l'obiettivo perseguito dal legislatore della riforma non potrebbe di certo essere realizzato in presenza un'interpretazione estensiva della nozione di causalità diretta<sup>32</sup>.

### **1.2.1 La causalità diretta tra volontà del legislatore e interpretazione della giurisprudenza**

Dal rapporto presentato al Senato, si desume che il legislatore della riforma del 2000 intendeva per causa diretta tanto la causalità immediata (prossimità nel tempo e nello spazio tra l'evento e la condotta), quanto quella che "normalmente" determina l'evento, rendendolo conseguenza "quasi automatica e quindi prevedibile", così facendo riferimento alla teoria della causalità adeguata<sup>33</sup>. L'effetto è stato quello di mettere sullo stesso piano due concezioni sensibilmente differenti della causalità<sup>34</sup>.

---

<sup>31</sup> Ma la distinzione vale solo per le persone fisiche e non anche in caso di responsabilità delle persone giuridiche, sul punto F. Desportes, F. Le Guehec, *op. cit.*, p. 409.

<sup>32</sup> Y. Mayaud, *obs. Cass. Crim. 29 oct. 2002, Bull. crim. n. 196, RSC 2003*, p. 330.

<sup>33</sup> « *Des exemples et analyses précédemment cités, il est possible de retenir que la cause directe est le plus souvent immédiate même si ce n'est pas toujours le cas, que cette cause directe est celle qui entraîne normalement ou nécessairement le dommage, celle dont le dommage est la conséquence quasiment automatique et donc prévisible* », *Rapport n° 177 (1999-2000) de M. Pierre FAUCHON, fait au nom de la commission des lois, déposé le 20 janvier 2000, www.senat.fr*

<sup>34</sup> B. Cotte, D. Guihal, *op. cit.*

La giurisprudenza, dal suo canto, non esita a fare ricorso alla teoria della causalità adeguata per stabilire la causalità diretta<sup>35</sup>, giungendo non di rado a interpretazioni estensive di tale tipo di causalità, con ciò frustrando lo scopo che il legislatore della riforma si era prefissato<sup>36</sup>. Così è accaduto quando, per stabilire il carattere diretto del nesso eziologico, la giurisprudenza ha fatto riferimento non solo alla causa immediata, ma anche alla condotta considerata “parametro determinate”, cioè quella che rileva nell’ambito di una dinamica causale che ha portato alla verifica dell’evento, nonostante la prima non sia cronologicamente prossima a quest’ultimo<sup>37</sup>, ovvero anche quando tale vicinanza cronologica sia presente, allorché l’agente non interviene a interrompere un decorso causale che già di per sé avrebbe portato alla realizzazione dell’evento<sup>38</sup>. Se ne ricava che la giurisprudenza tende a riconoscere una causalità diretta sia quando il comportamento consiste nell’assunzione di una decisione che porta all’esecuzione materiale della condotta da parte di terzi (come avviene nell’ambito della responsabilità medica allorché l’operazione decisa dal medico viene poi eseguita dai suoi sottoposti), sia quando vi sono più cause che concorrono alla determinazione dell’evento<sup>39</sup>, senza che si valorizzi la prossimità tra azione e offesa secondo il criterio della *causa proxima*, con la conseguenza che, in presenza di concause, è ritenuto diretto anche il contributo causale che invece presenta i caratteri della causalità indiretta.

### 1.2.2 Il nesso di causalità indiretto

La presenza di due o più concause pone le basi per l’accertamento della causalità indiretta, riferibile alla condotta di agenti mediati o indiretti. A mente del comma quarto

---

<sup>35</sup> Anche se c’è chi nota che la giurisprudenza non si preoccupa troppo di applicare una teoria piuttosto che un’altra, procedendo più “par intuition”, X. Pin, *Droit pénal général*, 12 ed., Paris, Dalloz, 2020, p. 198.

<sup>36</sup> *Cass. Crim. 25 sept. 2001, Bull. crim. n. 188, RSC 2002, p. 101, obs. Y. Mayaud e Cass. Crim. 29 oct. 2002, Bull. crim. n. 196, RSC 2003, p. 330, obs. Y. Mayaud, cit.*; entrambe le pronunce fanno riferimento al “paramètre déterminant” per stabilire l’efficacia causale diretta tra la condotta del reo (nel primo caso eccesso di velocità nella guida di un’auto e nel secondo un intervento di chirurgia estetica) e l’evento (rispettivamente, morte di un altro conducente e morte della paziente).

<sup>37</sup> Y. Mayaud, *RSC 2003, cit.*, p. 331; B. Cotte, D. Guihal, *op. cit.*

<sup>38</sup> *Cass. crim. 13 nov. 2002, bull. Crim. n. 203*, sulla morte di un neonato per ematoma cranico, in cui oltre al medico che aveva innescato il decorso causale che avrebbe poi portato alla morte, viene condannata sulla base di un nesso di causalità riconosciuto come diretto anche la pediatra che non aveva preso le misure idonee a evitare il verificarsi della morte. Sebbene in questo caso la giurisprudenza abbia fatto applicazione del criterio della *causa proxima*, la dottrina critica questa decisione, ritenendo che mentre la causalità è diretta con riferimento al primo degli imputati, sarebbe indiretta con riguardo alla condotta omissiva della pediatra, che si aggiunge alla prima azione senza però essere all’origine del processo mortale (Y. Mayaud, *RSC 2003, cit.*).

<sup>39</sup> B. Cotte, D. Guihal, *op. cit.*

dell'art. 121-3 c. pén, sono tali le persone fisiche che hanno creato o contribuito a creare la situazione che ha permesso la realizzazione del danno o che non hanno preso le misure idonee a evitarlo.

Una dottrina tenta di riassumere le conclusioni cui giunge la giurisprudenza affermando che è indiretta la causa ricollegabile al difetto di vigilanza di un soggetto posto a capo di un'impresa, di un'organizzazione o di un'attività di cui egli è responsabile, il cui mancato controllo ha contribuito a creare la situazione fattuale che ha reso possibile il verificarsi dell'evento, quand'anche la causa diretta sia ravvisabile nella condotta della stessa vittima, di un terzo o nel verificarsi di un evento naturale<sup>40</sup>. Altri<sup>41</sup> non mancano di evidenziare che talvolta la presenza di un evento naturale che si inserisce nella serie causale che porta alla verifica dell'evento non è in grado di declassare il contributo dell'agente da diretto a indiretto<sup>42</sup>. In particolare, la giurisprudenza è dell'opinione che la

---

<sup>40</sup> F. Desportes, F. Le Gunehec, *op. cit.*, p. 410. La giurisprudenza ha individuato la causa diretta nella condotta della vittima nel caso di uno studente che si prestava a un gioco mortale, ritenendo per contro indiretta la causa consistente nel difetto di vigilanza dell'insegnante, la quale veniva perciò assolta, non essendo accertata una *faute caractérisée* (Cass. crim. 10 déc. 2002, Bull. Crim. N. 223, che nota: « *l'institutrice, qui ignorait qu'il se livrait depuis peu au jeu dangereux dont il a été victime, n'a pas violé d'obligation particulière de prudence ou de sécurité prévue par la loi ou le règlement, les dispositions invoquées des lois du 18 septembre 1937 et du 10 juillet 1989 alors en vigueur ne prescrivant aucune obligation de cette nature, et qu'elle n'a pas commis de faute caractérisée exposant la victime à un risque qu'elle ne pouvait envisager* »); nella condotta del terzo che comportava la fuoriuscita d'acqua da una diga con conseguente morte degli escursionisti (Cass. crim. 18 juin 2002, Bull. Crim. n. 139, che confermava l'assoluzione delle insegnanti per il reato di omicidio e lesione degli allievi che venivano sorpresi durante un'escursione da una piena del fiume dovuta alla fuoriuscita d'acqua dagli sfioratori della diga. Afferma la corte: « *la seule cause directe des dommages est un lâcher d'eau effectué sans précaution par des préposés d'Electricité de France, relève, par les motifs repris aux moyens, que ni l'institutrice, qui avait obtenu de l'inspecteur de l'Education nationale l'autorisation de conduire ses élèves dans le lit du Drac pour leur faire découvrir l'environnement local avec l'assistance d'une animatrice qualifiée (...), ni la directrice, qui s'était bornée à émettre un avis favorable à la demande d'autorisation formulée par l'institutrice, n'ayant pu envisager le risque auquel étaient exposés les élèves, n'ont commis une faute caractérisée et n'ont violé de façon manifestement délibérée une obligation particulière de prudence ou de sécurité prévue par la loi ou le règlement ; Attendu qu'en statuant ainsi, la cour d'appel a fait l'exacte application des articles 121-3 et 221-6 du Code pénal, dans leur rédaction résultant de la loi du 10 juillet 2000*); nell'evento naturale nel caso di una slavina di neve (Cass. crim. 26 nov. 2002, Bull. Crim. n. 211, che confermava l'assoluzione dai reati di omicidio e lesioni colpose di alcuni studenti durante un'escursione con racchette da neve a carico del direttore di un centro sportivo che non era presente sul luogo al momento dei fatti, ritenendo che la sola causa diretta dell'evento era riconducibile alla slavina provocata dal passaggio del gruppo di escursionisti accompagnati dalla guida, la quale ultima era stata condannata), mentre la sua condotta rilevava come causa indiretta che richiede una *faute caractérisée* nella specie non accertata in quanto non vi era nessuna normativa applicabile alla pratica delle escursioni con racchette. Tutte le decisioni qui indicate sono commentate da Y. Mayaud, RSC 2003, pp. 330 ss.

<sup>41</sup> E. Dreyer, *La causalité directe de l'infraction*, Dr. Pén., 2007, n° 6, Juin 2007, étude 9.

<sup>42</sup> Come è accaduto in un arresto in cui la Corte di cassazione ha escluso la rilevanza del fattore naturale del passaggio improvviso di un cinghiale nella carreggiata che veniva investito da un automobilista che conduceva il proprio veicolo oltre i limiti stabiliti e provocava la morte di una persona, ritendo la condotta di costui come causa diretta dell'evento morte, (« *l'excès de vitesse commis par le prévenu est constitutif d'une faute en relation directe et certaine avec le décès* »), Cass. crim., 25 sept. 2001, Bull. crim. 2001, n° 188 ; RSC 2002, p. 101, obs. Y. Mayaud.

causalità diretta è interrotta solo laddove l'azione umana di un terzo si inserisca nella serie causale che porta alla verificazione dell'evento, mentre un simile effetto interruttivo del nesso causale non è riconosciuto al fattore naturale che, pur frapponendosi tra la condotta dell'agente e l'evento, lascia invariato il giudizio sul carattere diretto della causalità, col risultato -di cui si è poco prima detto- che la nozione di causa diretta viene estesa in modo eccessivo<sup>43</sup>. Ma a rigore, senza considerare l'interpretazione estensiva data dalla giurisprudenza, sarebbe più corretto dire che vi è causalità diretta solo quando la condotta è la causa unica o esclusiva, ovvero la causa immediata o determinante l'evento<sup>44</sup> e che, in tutti gli altri casi, vi è causalità indiretta.

Nondimeno, l'atteggiamento della giurisprudenza è letto osservando che in presenza di più fatti che si inseriscono nella catena causale, solo la condotta del terzo può porre le basi per il riconoscimento del contributo indiretto dell'agente principale perché esso solo costituisce un fatto penalmente rilevante e quindi meritevole di sanzione; non così per il fattore naturale o il fatto proprio della vittima, che costituiscono circostanze penalmente irrilevanti e, di conseguenza, giuridicamente inidonee a qualificare il comportamento dell'agente come causalmente indiretto<sup>45</sup>.

Nonostante le perplessità che si riscontrano in presenza di due o più concause, dallo studio della giurisprudenza emerge una tendenza costante a riconoscere la causalità indiretta a beneficio di soggetti che si trovano a capo di organizzazioni complesse in caso di errori di organizzazione o mancato controllo dei soggetti sottoposti. Ad esempio, è considerato agente indiretto o mediato il responsabile di un'impresa in caso di incidenti sul posto di lavoro<sup>46</sup> o di reati colposi commessi da terzi<sup>47</sup>, così come il sindaco di una città per i reati colposi posti a suo carico<sup>48</sup>. Ancora, vi è causalità indiretta nel caso di ispettori o soggetti titolari di posizioni di controllo, istruttori e insegnanti<sup>49</sup>. Va precisato

---

<sup>43</sup> E. Dreyer, *La causalité directe*, cit., par. 20.

<sup>44</sup> D. Commaret, *La loi du 10 juillet 2000 et sa mise en œuvre par la chambre criminelle de la Court de cassation*, *Gaz. Pal.*, 12-13 avril 2002, p. 4.

<sup>45</sup> E. Dreyer, *Droit pénal*, op. cit., p. 636.

<sup>46</sup> *Cass. crim. 16 janv. 2001, Bull. n. 14 e 15, sulla responsabilità dell'imprenditore per la morte di un dipendente che cadeva da un'impalcatura, senza che fosse dotata di dispositivi di protezione; cass. crim. 12 sept. 2000, Bull. n. 268; 24 oct. 2000, Bull. n. 308.*

<sup>47</sup> *Cass. Crim. 19 oct. 2004, n. 04-82.485, RSC 2005, p. 291, obs. Vermelle*, sulla responsabilità del capo di un'impresa che ometteva di adottare le misure idonee a evitare il rigetto di sostanze inquinanti, la cui causa diretta consisteva nell'intervento di operai dipendenti di un'impresa terza.

<sup>48</sup> *Cass. crim., 2 déc. 2003, n° 03-83.008, Bull. crim. 2003, n° 231* sulla responsabilità del sindaco per la morte di un bambino che giocava in un'area gioco ritenuta pericolosa dove decedeva per trauma al cranio.

<sup>49</sup> *Cass. crim., 20 mars 2001, n° 00-84.384, Bull. Crim. 2001 n. 75 p. 244*, che annulla con rinvio la sentenza della corte d'appello che aveva condannato l'insegnante per omicidio colposo con riferimento alla morte di un allievo che durante la ricreazione si lanciava dal terrazzo della scuola. Osserva la corte che l'omessa

che in molte di tali pronunce la corte si preoccupa di qualificare l'atteggiamento psicologico dell'imputato come *faute caractérisée* senza soffermarsi tanto sul piano causale, riconoscendo implicitamente la natura indiretta del contributo apportato.

In conclusione, dall'analisi della giurisprudenza emerge che i giudici francesi sono restii a riconoscere un legame eziologico indiretto nel caso di conducenti di veicoli e incidenti stradali e nell'ambito della responsabilità medica; mentre i soggetti aventi un compito gestionale o decisionale e posti a capo di organizzazioni complesse tanto pubbliche, quanto private, vedono riconosciuto il loro contributo indiretto, con la conseguenza che la riforma trova applicazione solo in questi casi, per i quali l'accusa dovrà dimostrare la *faute qualifiée*, essendo insufficiente la prova di una *faute simple*<sup>50</sup>.

### 1.2.3 Certezza del nesso di causalità

Naturalmente la causalità diretta, così come quella indiretta, si conciliano tanto con la condotta attiva, quanto con quella omissiva<sup>51</sup> e, in ogni caso, sia in presenza di un nesso di causalità diretto, che indiretto, è necessario che il legame eziologico sia stabilito in modo certo: *“la causalité est ou n'est pas”*<sup>52</sup>. La certezza del nesso di causalità è il diretto precipitato del principio di legalità e del suo corollario *« nul n'est responsable pénalement que de son propre fait »* (C. pén., art. 121-1)<sup>53</sup>, per questo la causalità penale, contrariamente a quella civile, non tollera interpretazioni estensive.

Così la causalità è stata esclusa laddove lo stato delle informazioni disponibili non consentiva di provare che la morte di alcuni avventori di una catena di ristoranti fosse riconducibile con certezza alla consumazione della carne lì servita piuttosto che all'ingerimento di altri alimenti da parte delle vittime<sup>54</sup> o ancora nell'ambito di un

---

vigilanza dell'insegnante è causa non diretta dell'evento che pertanto va valutata alla luce delle disposizioni più favorevoli introdotte dalla riforma del 10 luglio 2000.

<sup>50</sup> D. Commaret, *La responsabilité pénale des décideurs*, cit.

<sup>51</sup> In tema di omissione va precisato che nel codice penale francese non vi è una norma analoga a quella dell'art. 40 cpv. del codice penale italiano, di talché la punibilità delle condotte omissive corrispondenti alla categoria italiana del reato omissivo improprio avviene per mano della giurisprudenza attraverso la sola interpretazione della norma di parte speciale, sempreché questa presenti una descrizione del fatto generica che consenta di farvi rientrare tanto la condotta attiva, quanto quella omissiva. Si rinvia a X. Pin, *Droit pénal général*, 12 ed., Paris, Dalloz, 2020, pp. 185 e 186.

<sup>52</sup> E. Dreyer, *op. cit.*

<sup>53</sup> E. Dreyer, *op. cit.*

<sup>54</sup> Cass. 1 oct. 2003, *Bull. crim. n. 177*: *« rien ne permet de penser en l'état que la maladie de Laurence B... et Arnaud C... trouve son origine dans la consommation de viande servie par les restaurants Buffalo Grill plutôt que par les diverses autres sources alimentaires des victimes »*; seguiva la declaratoria di nullità del capo di imputazione per omicidio colposo.

procedimento a carico di due funzionari della direzione dipartimentale per gli affari sanitari e sociali che non avevano disposto l'ospedalizzazione di un soggetto che, in preda a un turbamento psichico che ne aveva abolito la capacità di discernimento, feriva a morte la vittima, non essendo dimostrato in modo certo che l'ospedalizzazione avrebbe impedito il fatto<sup>55</sup>. E, non da ultimo, il nesso causale non è stato dimostrato con certezza nel caso di morte da infezione da virus HIV contratto a causa di trasfusioni<sup>56</sup>.

Si tratta dei problemi che la dottrina e la giurisprudenza italiane affrontano facendo ricorso all'impiego delle leggi scientifiche di copertura per sopperire all'efficacia euristica limitata della teoria condizionalistica. Di tale prassi non vi è traccia nella giurisprudenza francese, né tantomeno la teoria delle leggi scientifiche è discussa in dottrina.

Talvolta la certezza del nesso causale si valuta con riferimento alla perdita di *chance*, costantemente richiamata nei casi di responsabilità medica per escludere l'imputazione di omicidio colposo a carico del personale sanitario allorché sia dimostrata "*la perte d'une chance de survie, et non la perte de toutes chances de survie*"<sup>57</sup>. In altri termini, occorre dimostrare la perdita di tutte le chance di sopravvivenza, piuttosto che la perdita di una sola chance, affinché la relazione causale tra la condotta e l'evento morte sia stabilita in modo certo.

In dottrina si fa riferimento alla *cause étrangère* per escludere la certezza del *lien de causalité*, affermando che il nesso causale non è certo se è provata la sopravvenienza di un avvenimento imprevedibile e invincibile, che corrisponda al caso fortuito, al fatto di un terzo o della stessa vittima<sup>58</sup>.

Concludendo, è possibile sintetizzare quanto sin qui osservato dicendo che il nesso non può ritenersi accertato allorché è impossibile stabilire l'origine dell'offesa, vi è la rottura del legame eziologico tra azione ed evento da parte di una *cause étrangère* che

---

<sup>55</sup> Cass. crim., 4 mars 2008, n° 07-81.108, *JurisData* n° 2008-043367 ; M. Véron, *La certitude du lien de causalité*, *Droit pénal* n° 6, Juin 2008, *comm.* 82.

<sup>56</sup> Crim. 18 juin 2003, Bull. crim. n° 127 ; Recueil Dalloz, 2005. Jur. 195, note A. Prothais ; RSC 2003. 781, obs. Y. Mayaud ; JCP 2003. II. 10121, note Rassat ; Dr. pénal 2003. Comm. 97, obs. Véron.

<sup>57</sup> Cass. Crim. 20 mars 1996, Bull. crim. n° 119 ; Cass. Crim. 22 mars 2005, Dr. pénal 2005. Comm. 103, obs. Véron ; Cass. Crim. 3 nov. 2010, Bull. Crim. n° 170. Per i casi in cui invece la perdita di tutte le chances di sopravvivenza è stata dimostrata, con conseguente condanna: Cass. Crim. 9 juin 1977, Bull. crim. n° 212.

<sup>58</sup> E. Dreyer, *Droit pénal*, *op. cit.*, p.626-627.

può consistere in un fatto sopravvenuto<sup>59</sup> o in una condizione preesistente (come lo stato di salute della vittima nei casi di responsabilità medica)<sup>60</sup>.

A questo punto è bene passare all'analisi dell'elemento soggettivo de *les infractions non intentionnelles*, cioè la *faute*.

### 1.3 La *faute*

Nella tripartizione del reato, i *crimes* sono sempre dolosi, le *contraventions* non richiedono un elemento soggettivo, salvo il regolamento preveda diversamente, mentre i *délits* possono essere sia dolosi che colposi.

Col termine *faute* si fa genericamente riferimento all'elemento psicologico del reato, che si distingue in *faute intentionnelle* e *faute non intentionnelle*. La prima coincide col dolo, quindi sono sempre *infractions intentionnelles* i *crimes*, che sono i reati più gravi. La *faute non intentionnelle* descrive in generale i reati colposi, all'interno dei quali il termine *faute* è impiegato anche per indicare la colpa. L'ordinamento francese conosce varie accezioni di colpa che, dopo la riforma del 10 juillet 2000 la dottrina raggruppa in due grandi categorie: *la faute d'imprudence ordinaire* o *faute simple*, come sopra è stata già definita, e due tipi di *fautes d'imprudence qualifiées*.

La colpa ordinaria postula la violazione di una regola di diligenza o di prudenza prevista da una regola di condotta, che può anche trovare fonte in massime di esperienza (*faute d'imprudence* e *de negligence*) e non solo in leggi o regolamenti (*manquement à une obligation de prudence ou de sécurité prévue par la loi ou le règlement*)<sup>61</sup>. Essa va accertata in concreto per espressa previsione legislativa (« *s'il est établi que l'auteur des faits n'a pas accompli les diligences normales compte tenu, le cas échéant, de la nature de ses missions ou de ses fonctions, de ses compétences ainsi que du pouvoir et des moyens dont il disposait* »)<sup>62</sup> e in presenza di un contributo *directo* nella causazione dell'evento, mentre non è sufficiente in caso di causalità indiretta. Quest'ultima, come si

---

<sup>59</sup> Come ad esempio l'infezione nosocomiale contratta dalla vittima di un incidente in ospedale, che decedeva proprio a causa dell'infezione, la quale è stata considerata fattore in grado di spezzare il nesso causale tra la condotta dell'automobilista e la vittima dell'incidente, *Cass. crim.*, 5 oct. 2004, *Bull. crim.* 2004, n° 230.

<sup>60</sup> B. Cotte, D. Guihal, *op. cit.*

<sup>61</sup> Secondo la giurisprudenza, la disattenzione del conducente alla guida di un veicolo a motore rappresenta un *manquement à une obligation de prudence prévue par le code de la route* (*Cass. crim.* 3 déc. 2019, n. 19-82492, *Dr. Pén.* 2020, obs. Ph Conte), ma la dottrina è critica sul punto, poiché osserva che in questo modo si alleggerisce il carico probatorio dell'accusa, X. Pin., *op. cit.*, p. 217.

<sup>62</sup> F. Desportes, F. Le Guehec, *op. cit.*, p. 456.

è già avuto modo di notare, incide sull'elemento soggettivo, qualificandolo: l'autore (persona fisica e non giuridica) di un comportamento che ha determinato il prodursi del danno (*auteur indirect*) o che non ha adottato le misure idonee a evitarlo (*auteur médiate*), e che con ciò ha solo *indirettamente* causato l'evento, è penalmente responsabile solo se è riscontrata una delle due *fautes qualifiées*, cioè, alternativamente, la *faute délibérée* o la *faute d'imprudence caractérisée*, entrambe previste dal comma 4 dell'art. 121-3.

La prima, cioè la *faute délibérée*, è quella di chi ha “*violé de façon manifestement délibérée une obligation particulière de prudence ou de sécurité prévue par la loi ou le règlement*”. La norma chiede come necessaria una violazione “cosciente” (in questo senso va intesa l'espressione “*manifestement délibéré*”<sup>63</sup>) di un obbligo di prudenza o sicurezza previsto dalla legge o dal regolamento. Sebbene l'espressione sia la stessa di quella adoperata al c.3 con riferimento alla *faute simple*, la dottrina<sup>64</sup> osserva che si tratta di un obbligo diverso, in quanto il legislatore ha premura di precisare – solo in questo caso e non anche nel c. 2 – che deve trattarsi di una *obligation particulière*, e ciò è stato correttamente inteso dalla giurisprudenza come un obbligo desumibile non da una prescrizione generale<sup>65</sup>, ma da una fonte (legge o regolamento e non altre) necessariamente scritta<sup>66</sup> che contenga un obbligo preciso, la quale, inoltre, deve essere indicata nella sentenza di condanna, pena l'assenza di base legale della stessa con conseguente assoluzione<sup>67</sup>. Queste caratteristiche, che fanno della *faute délibérée* una colpa difficile da accertare, non hanno sempre avuto il risultato di negare la responsabilità penale; piuttosto, non è raro che la giurisprudenza faccia ricorso al secondo tipo di *faute qualifiée*, cioè la *faute caractérisée*, come manifestazione soggettiva residuale, nella quale far convergere anche una *faute délibérée* che non può essere sanzionata per

---

<sup>63</sup> X. Pin., *op. cit.* p. 220, E. Dreyer, *p. cit.*, p. 711.

<sup>64</sup> X. Pin., *op. cit.* p. 220

<sup>65</sup> Cass. Crim. 25 juin, 1996, *Bull. crim. n. 274*, che conferma la sentenza di secondo grado la quale a sua volta confermava l'archiviazione per il reato “de mise en danger d'autrui” a carico dell'allora sindaco di Parigi per non aver preso misure idonee a evitare un incremento dell'inquinamento dell'aria, precisando che: « *l'article L. 131-2. 6° du Code des communes, alors en vigueur, qui confie au maire de façon générale le soin de prévenir et faire cesser tous les événements survenant sur le territoire de sa commune et de nature à compromettre la sécurité des personnes, ne crée pas à sa charge d'obligation particulière de sécurité au sens de l'article 223-1 du Code pénal (reato de mise en danger d'autrui), en raison du caractère général de ses prescriptions* ».

<sup>66</sup> E. Dreyer, *op. cit.*, p. 711, Cass. Crim., 10 déc. 2002, *Bull. Crim. 2002, n. 223*.

<sup>67</sup> Cass. Crim., 18 juin 2002, n. 01-86.539, *Recueil Dalloz*, obs. di F. Gauvin, p.240.

l'assenza di un obbligo particolare di prudenza o sicurezza previsto da leggi o regolamenti<sup>68</sup>.

Il risultato è che la *faute d'imprudence caractérisée* dissimula una *faute délibérée* "imperfetta"<sup>69</sup>, sebbene da essa differisca proprio poiché manca quella coscienza della violazione dell'obbligo cautelare, ma che si presta a tali manipolazioni da parte della giurisprudenza per via di una definizione meno rigorosa rintracciabile nella seconda parte del comma 4 dell'art. 121-3 c. pén.

È *faute caractérisée* la colpa « *qui exposait autrui à un risque d'une particulière gravité qu'elles ne pouvaient ignorer* » (art. 121-3, comma 4). Quest tipo di colpa si connota per la presenza di tre elementi: a) deve trattarsi di una *faute caractérisée*, cioè avente un carattere "bien marqué" e una "particulière intensité" che lascino apparire "un certain degré de gravité"<sup>70</sup>, sicché è esclusa la colpa lieve o lievissima; b) è necessario che si esponga un soggetto terzo a un rischio grave; c) tale rischio non può essere ignorato dall'agente, requisito inteso nel senso che non è verosimile che il soggetto non potesse averne contezza.

Quanto al rischio, è bene precisare che la sua gravità è accertata in astratto, a prescindere dalla verifica del danno finale che può anche rivelarsi insignificante<sup>71</sup>; quanto alla colpa, è bene ribadirlo, non è in alcun caso con previsione o coscienza: il rischio di particolare gravità non deve essere stato percepito dall'agente e il rimprovero gli è mosso proprio perché ha esposto altri a un rischio prevedibile<sup>72</sup> senza averne avuto contezza<sup>73</sup>.

Come è dato notare, i due tipi di *faute qualifiées* presentano delle differenze, poiché la *faute caractérisée* non presuppone una regola cautelare che sia necessariamente posta da leggi o regolamenti, potendo essa essere prevista anche da una fonte sotto-ordinata (ad es. una circolare) o da regole di fonte sociale, né pretende una violazione manifestamente deliberata, cioè coscienza, di tale regola. Nondimeno, esige la prevedibilità dell'esistenza

---

<sup>68</sup> Come è avvenuto nel caso di violazione "cosciente" di una regola cautelare contenuta in una legge che non era stata regolarmente promulgata nella Polinesia francese, *Cass. crim. 15 oct. 2002, Bull. Crim. 2002, n. 186; Dr. Pén. 2003, comm. 4, obs. M. Véron.*

<sup>69</sup> E. Dreyer, *Droit pénal, cit.*, p. 703.

<sup>70</sup> F. Desportes, F. Le Guehec, *op. cit.*, p. 470.

<sup>71</sup> E. Dreyer, *op. cit.*, p. 702.

<sup>72</sup> È ad esempio prevedibile il rischio di una patologia da parte del medico che ha ommesso di porre ulteriori domande al paziente, il quale decedeva a causa di un infarto, *Cass. crim. 2 déc. 2003, Bull. Crim 2003, n. 226, AJ Pén. 2004, p. 118.*

<sup>73</sup> *Cass. Crim, 9 avr. 2019, Dr. Pén. 2019, comm. 103, obs. Ph. Conte ; Cass. Crim, 26 juin 2001, Dr. Pén. 2001, comm. 124, obs. M. Véron.*

del rischio e suppone in ogni caso una certa gravità della colpa e ciò supera l'idea che la *faute caractérisée* sia il grado di colpa più lieve e che la *faute délibérée*. Si tratta, piuttosto, di tipi di colpa che si situano su piani leggermente differenti, ma accomunati dal presupposto della causalità indiretta. Come è stato efficacemente notato, dopo la legge del 10 luglio 2000 è richiesta una gravità della colpa inversamente proporzionale alla causalità<sup>74</sup>: mentre il contributo causale diretto richiede la colpa semplice, la causalità indiretta non è penalmente rilevante se non sia provata una colpa qualificata, nelle due forme di cui si è appena detto, con un significativo restringimento dell'ambito applicativo dello strumento penale, che tuttavia non ha portato alla depenalizzazione auspicata.

Va in ogni caso ricordato che la distinzione tra *faute simple* e *fautes qualifiées* serve per rendere incriminabile il comportamento causale indiretto, ma non ad attenuarne il trattamento sanzionatorio: esso rimane identico sia per la colpa ordinaria che per quella qualificata, fatta eccezione per alcuni reati di parte speciale (come l'omicidio colposo, art. 221-6 e l'omicidio stradale, art. 221-6-1 c.2) che prevedono la *faute délibérée* come circostanza aggravante<sup>75</sup>.

### 1.3.1 L'elemento morale nelle contravvenzioni

Se con riguardo ai reati rientranti nelle categorie dei *crimes* e dei *délits* sono le problematiche relative all'accertamento del contributo causale (e la relazione di quest'ultimo con la colpa) a costituire lo snodo critico della disciplina in materia di reati colposi, diverse sono le questioni che si pongono per la categoria delle contravvenzioni. Queste ultime presentano delle peculiarità che a tratti le allontanano dalla materia penale, tanto è vero che la dottrina mette in evidenza il loro carattere "para-penale"<sup>76</sup>.

Le differenze con i *crimes* e i *délits* sono molteplici e vanno oltre la fonte normativa (legislativa per i primi, regolamentare per i secondi), poiché coinvolgono anche la struttura degli illeciti contravvenzionali, caratterizzati dalla "*faiblesse de leur élément moral*": spesso, in assenza di disposizione contraria, non è necessario dimostrare la colpevolezza che è invece presunta dall'accertamento dell'elemento materiale

---

<sup>74</sup> F. Desportes, F. Le Guehec, *op. cit.*, p.473.

<sup>75</sup> Per questo motivo in dottrina c'è chi segue una classificazione diversa da quella in questa sede proposta, che segue non il legame tra causalità (diretta e indiretta) e colpa, ma il trattamento sanzionatorio, così distinguendo *fautes simples* e *aggravé*. Rientra in quest'ultima categoria la *faute délibérée*, mentre gli altri due tipi di colpa sarebbero di tipo semplice. E. Dreyer, *op. cit.*, p. 714.

<sup>76</sup> E. Dreyer, *Droit pénal, cit.*, p. 81.

dell'infrazione<sup>77</sup>. La possibilità di un'imputazione su base oggettiva di un reato contravvenzionale emerge dall'ultimo comma dell'art. 121-3 c. pén, a mente del quale “*il n'y a pas de contravention en cas de force majeure*”; più in generale, l'intero dibattito circa l'elemento soggettivo delle contravvenzioni risente dell'assenza di un riconoscimento costituzionale del principio di colpevolezza<sup>78</sup>. Così le perplessità sull'elemento morale delle contravvenzioni sono spesso state risolte affermando che in generale le contravvenzioni non richiedono la prova dell'elemento soggettivo, salvo che la legge disponga diversamente<sup>79</sup>, per cui è sufficiente la sola prova dell'elemento materiale della condotta<sup>80</sup>. Lo stesso Conseil Constitutionnel ha avuto modo di affermare che in materia contravvenzionale la presunzione di colpa non stride con la presunzione di innocenza di cui all'art. 9 della Dichiarazione del 1789 laddove sia ammessa la prova contraria<sup>81</sup>.

Inoltre, delle peculiarità rispetto agli altri reati si riscontrano sul piano processuale. Le contravvenzioni, infatti, vengono accertate all'esito di un procedimento semplificato dinanzi al tribunal de police<sup>82</sup>, in assenza di contraddittorio preventivo<sup>83</sup>, ovvero con la procedura dell'ammenda forfetaria che prevede il pagamento diretto dell'importo della sanzione<sup>84</sup>.

Le peculiarità della categoria delle contravvenzioni sono alla base della tesi bipartita delle *infractions*, secondo la quale bisognerebbe distinguere tra *crimes* e *délits*, da un lato,

---

<sup>77</sup> E. Dreyer, *op. cit.*, p. 81.

<sup>78</sup> R. Sicurella, *cit.*, pp. 974 e 975.

<sup>79</sup> Come avviene ad esempio per la contravvenzione di cui all'art. R. 625 c. pén. sulle lesioni che comportino una incapacità lavorativa inferiore o pari a otto giorni, che sono “*violences volontaires*”, F. Desportes, F. Le Guehec, *op. cit.*, p. 428.

<sup>80</sup> Y. Mayaud, *Droit pénal général, cit.*, p. 352.

<sup>81</sup> Si tratta della decision n. 99-411 del 16 giugno 1999 del Cons. cost. ([www.conseil-constitutionnel.fr](http://www.conseil-constitutionnel.fr)), che sarà analizzata più avanti poiché avente ad oggetto contravvenzioni previste dal codice della strada francese.

<sup>82</sup> La composizione di questo tribunal è così stabilita dall'art. 523 c. proc. pén : « *Le tribunal de police est constitué par un juge du tribunal judiciaire, un officier du ministère public ainsi qu'il est dit aux articles 45 et suivants, et un greffier* ».

<sup>83</sup> Il contraddittorio può essere però disposto dal giudice ove ritenuto utile, come previsto dall'art. 525 commi 2 e 3 : « *Le juge statue sans débat préalable par une ordonnance pénale portant soit relaxe, soit condamnation à une amende ainsi que, le cas échéant, à une ou plusieurs des peines complémentaires encourues.*

*S'il estime qu'un débat contradictoire est utile, le juge renvoie le dossier au ministère public aux fins de poursuite dans les formes de la procédure ordinaire* ».

<sup>84</sup> Si rinvia agli artt. 524 – 528-2 c. proc. pén. sulla procedura semplificata e agli artt. 529 – 530-6 per la procedura dell'ammenda forfetaria.

e *contraventions*, dall'altro, le quali ultime resterebbero sempre nell'orbita della materia penale<sup>85</sup>.

Una soluzione differente è quella della depenalizzazione della categoria delle contravvenzioni, stante la loro natura di sanzioni para-penali, ma la dottrina avverte che si tratterebbe di una riforma non semplice: infatti, mentre da un lato la debolezza dell'elemento morale non è riscontrabile in tutte le contravvenzioni, dall'altro vi sono alcuni *délits* che rispondono a una logica disciplinare, cosicché per rendere coerente la risposta punitiva di tipo penale non basterebbe sopprimere la categoria delle contravvenzioni, ma sarebbe necessaria una valutazione che involge singole contravvenzioni e singoli delitti<sup>86</sup>. Proprio per evitare di ricorrere a una riforma di tale portata, il legislatore francese ha preferito intervenire sulla procedura per l'accertamento delle contravvenzioni, semplificandola, mentre ha mantenuto ferma la loro natura penale.

Queste brevi osservazioni in tema di *faute contraventionnelle* ci torneranno utili nel prosieguo, posto che molte sanzioni previste dal codice della strada hanno natura contravvenzionale, mentre nel sistema italiano sono sanzioni amministrative; tra questi figurano ad esempio l'eccesso di velocità e la sosta in zona vietata.

## **2. Responsabilità penale e veicoli a guida autonoma. Alcune riflessioni sulla recente riforma del Code de la route**

### **2.1 Code de la route e code pénal**

Il codice della strada francese, in vigore dal 1 giugno 2001, è frutto di un intervento ricognitivo, con il quale si è voluto raggruppare in un unico testo l'insieme delle norme relative alla circolazione stradale, portato a compimento con l'*ordonnance* del 22 settembre 2000.

Come anticipato, il codice si divide in due parti, quella legislativa e quella regolamentare. L'art. L. 121-1 del code de la route (c. route) apre il capitolo I del Tit. II del libro I, capitolo dedicato alla responsabilità penale<sup>87</sup>. In quanto responsabilità di

---

<sup>85</sup> È la tesi esposta da Y. Mayaud, *Droit pénal général*, Paris, PUF, 2018, pp. 350 ss.

<sup>86</sup> E. Dreyer, *Droit pénal*, cit., p. 81.

<sup>87</sup> Quanto alla numerazione, si seguono i criteri impiegati per i codici moderni, per cui la prima cifra corrisponde al libro in cui la norma è contenuta, la seconda cifra indica il titolo, la terza al capitolo e la quarta, preceduta da un trattino, è il numero che indica esattamente quella disposizione. Così, ad esempio, l'art. L121-1 si colloca nella parte di fonte legislativa del codice, nella specie nel libro primo, titolo secondo, capitolo primo ed è la disposizione d'apertura del suddetto capitolo.

natura penale, valgono per essa le disposizioni del codice penale; tale convergenza dei due testi legislativi verso la stessa materia ci richiede allora di analizzare i rapporti tra il codice della strada e il codice penale.

Le norme del code de la route sono regole proprie della materia della circolazione stradale, quindi speciali rispetto alle regole generali in tema di responsabilità penale, le quali continuano in ogni caso a trovare applicazione in questa materia, anche in assenza di un richiamo espresso<sup>88</sup>. Il codice penale, così come quello civile, sono infatti dei “codes majeurs”, la cui portata generale non fa eccezione in materia di circolazione stradale, cosicché il code de la route è definito come “code suiveur”, cioè codice “seguace”<sup>89</sup>. E in effetti, il codice penale francese contiene delle disposizioni peculiari applicabili esclusivamente alla circolazione stradale, introdotte dalla legge del 12 giugno 2003, tra cui i reati di omicidio e lesioni colpose stradali di cui agli artt. 221-6-1, 222-19-1 e 222-20-1 c.p., che sono espressamente richiamate dal code de la route<sup>90</sup>, secondo la pratica tipica del “code suiveur”<sup>91</sup>. Ciononostante i reati commessi durante la guida, sia pure non richiamati espressamente dal code de la route ovvero non formulati nel code pénal con riferimento espresso all’atto della guida, sono comunque applicabili in caso di circolazione stradale. È questo, ad esempio, il caso del *délit de risque causés à autrui*, previsto dall’art. 223-1 c. pén., così come molte contravvenzioni non richiamate dal codice della strada<sup>92</sup>.

Le sanzioni contenute nel code de la route operano invece solo in caso di circolazione stradale e, laddove vi sia un contrasto tra queste disposizioni e quelle del codice penale, devono ritenersi prevalenti le seconde<sup>93</sup>. Conseguentemente, tutte le disposizioni del code pénal in tema di responsabilità contenute nel libro II trovano applicazione anche per i reati commessi nell’ambito della circolazione stradale e, in particolare, è di certa applicazione l’art. 121-3 c. pén. sulla responsabilità colposa<sup>94</sup> sopra analizzato.

---

<sup>88</sup> L. Desessard, *Commentaires au Code de la route, art. 121-1*, Dalloz, 17 ed., p. 12.

<sup>89</sup> L. Desessard, *op. cit.*, p. 1122. Si tratta di una pratica legislativa, definita “*du code suiveur*”, che consiste nell’introdurre in un codice le disposizioni di altri codici che vengono non solo richiamate, ma anche interamente riportate nel codice “seguace” o minore.

<sup>90</sup> In particolare, i richiami si trovano nel titolo II del libro II che disciplina il “comportamento del conducente”, tra questi, si vedano l’art. L. 232-1, che richiama espressamente gli artt. 221-6-1 e 221-8 c. pén.; l’art. L. 232-2 che riproduce gli artt. 222-19-1, 222-20-1 e 222-44 c. pén. È anche richiamato il reato di omissione di soccorso (*délit de fuite*) di cui all’art. 434-10 e 434-45 c. pén., cui fa espressamente rinvio l’art. L. 231-1 code de la route.

<sup>91</sup> L. Desessard, *op. cit.*, p. 4.

<sup>92</sup> *Ibidem*, pp. 1122-23.

<sup>93</sup> *Ibidem*, p. 12.

<sup>94</sup> *Ibidem*, p. 13.

Vale a questo punto ricordare, seppur in modo sintetico, l'approccio della giurisprudenza in tema di colpa nell'ambito della circolazione stradale, che difficilmente tende a riconoscere un nesso di causalità indiretto (che comporta la prova della *faute caractérisée*) tra la violazione della regola sulla circolazione stradale e la morte della vittima<sup>95</sup> nonostante l'evento non sia cronologicamente prossimo alla condotta. Tale atteggiamento ermeneutico è tanto più evidente in caso di condotta integrante un eccesso di velocità, ove si sostiene che vi sia causalità diretta non solo quando questa è stata la causa immediata dell'offesa arrecata alla vittima, ma anche quando tale eccesso rappresenta uno degli elementi della serie causale che ha determinato l'evento<sup>96</sup>.

È in questo contesto che si inserisce la riforma del 2021 in tema di veicoli a conduzione autonoma.

### **3. La riforma del 2021 e la disciplina della responsabilità penale in caso di utilizzo di auto a guida autonoma**

Sul tema della responsabilità penale da circolazione su strada, il legislatore francese è intervenuto di recente per disciplinare l'utilizzo di auto a guida autonoma. Con la *loi d'habilitation* n. 2019-1428 del 24 dicembre 2019 (*loi d'orientation des mobilités*) il Parlamento francese ha autorizzato il Governo ad adottare una *ordonnance* al fine di adeguare le disposizioni del codice della strada alla circolazione su strada di veicoli dotati di sistema di guida delegata e di definire il relativo regime di responsabilità (art. 31 Loi n. 2019-1428)<sup>97</sup>.

---

<sup>95</sup> *Ibidem*, p. 14.

<sup>96</sup> *Cass. Crim. 25 sept. 2001, Bull. crim. n. 188*, in *Revue de science criminelle et de droit pénal comparé (RSC)*, 2002, p. 101, con osservazioni di Y. Mayaud. Ma naturalmente è richiesto che tale nesso di causalità sia certo, come è stato affermato dalla Cour de Cassation in un arresto giurisprudenziale relativo all'investimento di un pedone, deceduto qualche tempo dopo a causa di un'infezione. In quell'occasione la sentenza veniva cassata con rinvio poiché i giudici d'appello non avevano accertato il legame eziologico tra l'incidente e la morte per infezione (*Cass. crim. 5 oct. 2004, Bull. Crim. n. 230; RSC 2005*, p. 71, con osservazioni di Y. Mayaud).

<sup>97</sup> Art. 31 - *Dans les conditions prévues à l'article 38 de la Constitution, le Gouvernement est autorisé à prendre par voie d'ordonnance, dans un délai de vingt-quatre mois à compter de la promulgation de la présente loi, toute mesure relevant du domaine de la loi afin d'adapter la législation, notamment le code de la route, au cas de la circulation sur la voie publique de véhicules terrestres à moteur dont les fonctions de conduite sont, dans des conditions, notamment de temps, de lieu et de circulation, prédéfinies, déléguées partiellement ou totalement à un système de conduite automatisé, notamment en définissant le régime de responsabilité applicable.*

*Il peut être prévu à ce titre d'imposer la fourniture d'une information ou d'une formation appropriée, préalablement à la mise à disposition des véhicules à délégation de conduite, lors de la vente ou de la location de tels véhicules.*

In seguito, il Governo ha adottato l'*ordonnance* del 14 avril 2021 n. 2021-443, con la quale è stato inserito un terzo capitolo all'interno del Titre 2 del code de la route, rubricato "*Responsabilité pénale applicable en cas de circulation d'un véhicule à délégation de conduite (articles L123-1 à L123-4)*"<sup>98</sup>. Si tratta in sostanza dell'introduzione di quattro articoli, dei quali, in particolare: l'art. 123-1 c. route si occupa della responsabilità del conducente; l'art. 123-2 c. route della responsabilità del produttore; l'art. 123-3 c. route concerne l'accesso ai dati contenuti nei dispositivi di registrazione del sistema di guida autonoma ai fini delle indagini nell'ambito del processo penale; infine, l'art. 123-4 c. route rimanda a un decreto del Consiglio di Stato la determinazione delle modalità di applicazione del capitolo di nuova introduzione.

È poi inserito il Capitolo IX al libro III del code de la route, rubricato "*Dispositions applicables au véhicule à délégation de conduite*".

Strumento normativo di derivazione governativa che conosce un contributo preventivo e successivo del Parlamento, le *ordonnances* entrano in vigore dal giorno successivo alla pubblicazione nel Journal officiel, avvenuta nel caso di specie in data 15 aprile 2021<sup>99</sup>. Pertanto è a far data dal 16 aprile 2021 che il codice della strada francese conosce le nuove disposizioni che disciplinano un progresso tecnologico potenziale, ma non in atto. Infatti, non sono ancora in circolazione auto a guida autonoma di livello 3, cioè il livello che è autorizzato da questa riforma, poiché non ancora immesse in commercio.

Le *ordonnances* devono però essere ratificate entro il termine previsto dalla *loi d'habilitation*: in proposito l'art. 31 della loi n. 2019-1428 stabiliva che un progetto di legge di ratifica avrebbe dovuto essere depositato entro tre mesi dall'entrata in vigore

---

*Un projet de loi de ratification est déposé devant le Parlement dans un délai de trois mois à compter de la publication de l'ordonnance.*

<sup>98</sup> Per una breve sintesi del testo, si rinvia al *Rapport au Président de la République relatif à l'ordonnance no 2021-443 du 14 avril 2021 relative au régime de responsabilité pénale applicable en cas de circulation d'un véhicule à délégation de conduite et à ses conditions d'utilisation*, Journal Officiel, 15 avril 2021; in dottrina H. Christodoulou, *Quand la responsabilité pénale embarque à bord d'un véhicule à délégation de conduite*, Gaz. Pal. n°24, 29 juin 2021, p. 10 ; S. Detraz, *De la mauvaise conduite des véhicules autonomes en droit pénal*, Recueil Dalloz, 2021, p. 1039.

<sup>99</sup> Si tratta di provvedimenti adottati dal Governo in applicazione di una *loi d'habilitation* votata dal Parlamento, che fissa altresì il termine entro il quale dovrà procedersi alla ratifica dell'*ordonnance* da parte del Parlamento, come previsto dall'art. 38 Cost. Tale potere si esplica in materie che normalmente sarebbero di competenza legislativa. Si tratta di uno strumento per il quale sorgono delle tensioni col principio di riserva di legge, così come nel nostro ordinamento lo stesso problema è posto dall'uso del decreto legislativo e del decreto legge in materia penale.

dell'*ordonnance*. Tale progetto di ratifica è stato presentato il 23 giugno 2021 in Senato<sup>100</sup> e ciò sembrerebbe sufficiente a fugare possibili dubbi sul valore di legge del testo in esame, almeno alla luce della più recente giurisprudenza del Conseil constitutionnel sulle *ordonnances* non ratificate<sup>101</sup>. Un successivo decreto del 29 giugno 2021 è stato adottato dal Governo in applicazione dell'*ordonnance*, specificandone in parte i contenuti.

L'*ordonnance* del 14 aprile 2021, quindi, è un testo già in vigore, ciononostante non ha trovato ancora applicazione perché allo stato attuale non vi sono veicoli a guida autonoma in circolazione. Il loro impiego è circoscritto ai contesti di sperimentazione e per questo ricade entro la disciplina dell'*ordonnance* n. 2016-1057 del 3 agosto 2016 di cui si è accennato in premessa. Presumibilmente dovrà attendersi la conclusione delle

---

<sup>100</sup> Texte n° 710 (2020-2021) de Mme Barbara POMPILI, ministre de la transition écologique et M. Jean-Baptiste DJEBBARI, ministre délégué auprès de la ministre de la transition écologique, chargé des transports, déposé au Sénat le 23 juin 2021, Projet de loi ratifiant diverses *ordonnances* prises sur le fondement de la loi n° 2019-1428 du 24 décembre 2019 d'orientation des mobilités, art. 6.

<sup>101</sup> Posto che l'art. 38 Cost. fr. stabilisce che le *ordonnances* entrano in vigore a partire dalla loro pubblicazione, ma perdono efficacia se non vengono ratificate in modo espresso dal Parlamento entro il termine stabilito dalla *loi d'habilitation*, una recente decisione del Conseil constitutionnel (Conseil constitutionnel, 28 mai 2020, n° 2020-843 QPC, [www.conseil-constitutionnel.fr](http://www.conseil-constitutionnel.fr)) è intervenuta a precisare la natura giuridica delle *ordonnances* non ratificate. La pronuncia riguarda il caso di una *ordonnance* per la quale era già stato depositato il progetto di ratifica nei termini previsti dalla *loi d'habilitation*, ciononostante il Parlamento non aveva proceduto alla ratifica. Ebbene, il Conseil constitutionnel ha avuto modo di precisare che spirato il termine stabilito dalla *loi d'habilitation* per la ratifica, a mente dell'art. 38 Cost. fr., le disposizioni contenute nell'*ordonnance* possono essere modificate solo dalla legge qualora interessino materie di competenza legislativa. Da ciò la pronuncia fa discendere che, a far data dalla scadenza del termine per la ratifica, il contenuto dell'*ordonnance* acquista valore legislativo. Si legge al par. 11 della summenzionata sentenza: « si un projet de loi de ratification de l'ordonnance du 5 août 2013 a été déposé dans le délai fixé par l'article 12 de la loi du 27 décembre 2012, le Parlement ne s'est pas prononcé sur cette ratification. Toutefois, conformément au dernier alinéa de l'article 38 de la Constitution, à l'expiration du délai de l'habilitation fixé par le même article 12, c'est-à-dire à partir du 1er septembre 2013, les dispositions de cette ordonnance ne pouvaient plus être modifiées que par la loi dans les matières qui sont du domaine législatif. Dès lors, à compter de cette date, elles doivent être regardées comme des dispositions législatives ». In dottrina, sul punto, M. Verpeaux, *Actes administratifs - Ordonnances = dispositions législatives. A propos de la décision du Conseil constitutionnel du 28 mai 2020*, *La Semaine juridique. Administrations et collectivités territoriales*, 22 juin 2020, n° 25, p. 4; A. Levade, *Conseil constitutionnel et ordonnances : l'in vraisemblable revirement !*, *La Semaine juridique. Édition générale*, 29 juin 2020, n° 26, p. 1185 – 1186.

La vigenza delle *ordonnances* non ratificate per le quali sia stato presentato un progetto di ratifica non approvato è stata ribadita in una successiva pronuncia (Conseil constitutionnel, 3 juillet 2020, n° 2020-851/852 QPC, [www.conseil-constitutionnel.fr](http://www.conseil-constitutionnel.fr)) in cui il Conseil constitutionnel ha altresì affermato che le *ordonnances* possono essere oggetto di una questione di legittimità costituzionale (a conferma della loro assimilazione alla legge) e tale sindacato può persino estendersi anche alla *loi d'habilitation*. Per un commento alla decisione, si rinvia a A. Levade, *Ordonnance et loi d'habilitation, le Conseil constitutionnel étend son empire*, *La semaine juridique - édition générale - n° 30-35 - 27 juillet 2020*. Va osservato che nella prassi è raro che il Parlamento, in sede di ratifica, apporti delle modifiche al testo dell'*ordonnance* tramite emendamenti. La ratifica si riduce così a una mera formalità e ciò probabilmente attenua la portata della decisione del Conseil constitutionnel tesa a equiparare le *ordonnances* non ratificate (ma per le quali sia stata depositato un progetto di ratifica) alla legge.

sperimentazioni sulle auto a guida autonoma prima che esse inizino a circolare e che si creino i presupposti fattuali affinché la riforma del 2021 trovi applicazione.

Nel presente lavoro analizzeremo i primi due articoli introdotti dall'*ordonnance* del 14 aprile 2021, cioè gli artt. L 123-1 c. route e L 123-2 c. route.

#### **4. La responsabilità del conducente**

In generale, nel sistema francese la condotta diligente richiesta al conducente di un qualsivoglia veicolo a motore ha fonte normativa e, più specificamente regolamentare, in quanto è posta dall'art. R. 412-6 c. route, che dispone: « *Tout véhicule en mouvement ou tout ensemble de véhicules en mouvement doit avoir un conducteur. Celui-ci doit, à tout moment, adopter un comportement prudent et respectueux envers les autres usagers des voies ouvertes à la circulation. Il doit notamment faire preuve d'une prudence accrue à l'égard des usagers les plus vulnérables* ». La condotta negligente del conducente è allora penalmente rilevante alla luce del combinato disposto degli artt. 121-3 comma 3 c. pén (responsabilità per violazione di una "*obligation de prudence ou de sécurité prévue par la loi ou le règlement*") e R412-6 code de la route, che riempie di significato il comma terzo dell'art. 121-3 c. pén.

Con la riforma del 2021 si è contribuito a specificare l'obbligo di diligenza richiesto al conducente, isolando all'interno della generica categoria del 'conducente di un veicolo in movimento' quella del conducente di un veicolo dotato di un sistema di guida autonoma (*système de conduite automatisée (SCA)*), cui è dedicato l'art. 123-1 c. route. Il rapporto tra l'art. R412-6 c. route e l'art. L123-1 c. route sembra essere di specialità, ne senso che, mentre il primo impone un dovere di diligenza a qualsivoglia conducente di un veicolo a motore, l'art. L123-1 c. route riguarda i soli sistemi a guida autonoma, specificando quali debbano essere le accortezze richieste a chi conferisce al sistema di IA il controllo dinamico e quindi la guida.

L'art. 123-1 c. route assume come presupposto il primo comma dell'art. 121-1 c. route, che apre il Capitolo del code de la route dedicato alla responsabilità penale. In particolare, l'art. 121-1 c. route dispone: « *Le conducteur d'un véhicule est responsable pénalement des infractions commises par lui dans la conduite dudit véhicule* ».

Tale presupposto normativo viene utilizzato, da un lato, per escluderne l'applicazione e, dall'altro, per individuare i casi in cui risulta operante, così delimitando il perimetro di valenza di tale responsabilità alla persona del conducente del veicolo a guida autonoma. L'art. 123-1 c. route, pertanto, si presenta come una norma avente una portata negativa e una positiva: nel primo senso, esso esclude l'applicazione dell'art. 121-1 c. route e la responsabilità ivi prevista; nel secondo senso, esso invece la riafferma.

Cominciamo con l'analisi dell'esclusione della responsabilità.

L'art. 123-1 c. route afferma: « *Les dispositions du premier alinéa de l'article L. 121-1 ne sont pas applicables au conducteur, pour les infractions résultant d'une manœuvre d'un véhicule dont les fonctions de conduite sont déléguées à un système de conduite automatisé, lorsque ce système exerce, au moment des faits et dans les conditions prévues au I de l'article L. 319-3, le contrôle dynamique du véhicule* ».

La norma esclude la responsabilità del conducente per il caso in cui la condotta illecita sia posta in essere quando il sistema del pilota automatico eserciti “il controllo dinamico del veicolo”. In questo modo la nozione di “controllo dinamico” acquista un ruolo centrale nell'ambito dell'*ordonnance*, ma la stessa non ne fornisce una definizione, che invece è da rinvenire nel decreto del 29 giugno 2021, all'art. 2, ove si legge che è da intendersi come controllo dinamico la « *exécution de toutes les fonctions opérationnelles et tactiques en temps réel nécessaires au déplacement du véhicule. Il s'agit notamment du contrôle du déplacement latéral et longitudinal du véhicule, de la surveillance de l'environnement routier, des réactions aux événements survenant dans la circulation routière et de la préparation et du signalement des manœuvres* »<sup>102</sup>.

La norma sembra avere carattere ricognitivo, poiché il fatto che la responsabilità del conducente è esclusa allorché sia il sistema di IA ad avere il “controllo dinamico” della vettura è una conclusione cui si poteva giungere attraverso i principi generali del diritto penale. E in effetti, in tali casi, l'infrazione non può essere causalmente imputata alla condotta di chi è alla guida, atteso che con l'attivazione del sistema di guida autonoma egli diviene un “non conducente”<sup>103</sup>. E allora, la portata innovatrice della norma va ricercata altrove e sembra coincidere in particolare con la parte in cui si chiarisce che il

---

<sup>102</sup> In sostanza, il controllo dinamico consiste nel dominio degli spostamenti del veicolo (sia laterali che in avanti) oltreché della sorveglianza della strada, della reazione a eventi improvvisi e della preparazione e segnalazione delle manovre.

<sup>103</sup> S. Detraz, *op. cit.*

controllo dinamico può essere assunto dal sistema in presenza delle condizioni di cui all'art. 319-3 c. route, anch'esso introdotto con la riforma del 2021. Tale disposizione prevede che è il conducente a scegliere di attivare il sistema di guida autonoma, ma siffatta decisione deve essere assunta una volta che il sistema abbia informato il conducente della capacità di esercitare un controllo dinamico della vettura.

Quest'ultima disposizione si riferisce più al produttore che al conducente, in quanto sembra imporre che ogni attivazione del sistema di pilota automatico sia preceduta da un messaggio di avviso delle capacità del sistema rivolta al conducente, chiedendo quindi al produttore di impostare tale messaggio e di programmarne la comunicazione ogni volta che del sistema si faccia uso.

Sulla scia del comma 1, anche il comma secondo dell'art. 319-3 c. route indica delle prescrizioni rivolte al produttore, precisando che il sistema deve sempre avvertire il conducente allorché non sia più in grado di esercitare il controllo dinamico sulla vettura. In questo caso la legge impone che il sistema: i) avvisi il conducente; ii) chieda che il conducente riprenda il controllo della vettura; iii) inizi ed esegua una manovra con rischio minimo se il controllo non viene ripreso alla fine del periodo di transizione o in caso di un guasto grave<sup>104</sup>.

In sostanza, la responsabilità del conducente prevista dall'art. 121-1 c. route è esclusa quando egli abbia impostato la modalità di pilota automatico, dopo essere stato avvisato dal sistema che lo stesso assumerà il controllo dinamico della vettura.

In ogni caso, una volta scelta la modalità di guida autonoma il conducente non potrà rinunciare a qualsiasi forma di controllo sul sistema, poiché dovrà sempre essere pronto a riprendere il dominio della vettura a fronte di una domanda in tal senso da parte del sistema. Questa regola di diligenza è prescritta dallo stesso art. 123-1 c. route, che fonda l'idea di una guida "inversamente assistita", più che autonoma, poiché vi deve essere sempre una sorveglianza dell'uomo sul sistema, il quale non può essere lasciato solo nella

---

<sup>104</sup> Si riporta per comodità il testo dell'articolo in esame. *Article L319-3 : I - La décision d'activer un système de conduite automatisé est prise par le conducteur, préalablement informé par le système que ce dernier est en capacité d'exercer le contrôle dynamique du véhicule conformément à ses conditions d'utilisation.*

*II.- Lorsque son état de fonctionnement ne lui permet plus d'exercer le contrôle dynamique du véhicule ou dès lors que les conditions d'utilisation ne sont plus remplies ou qu'il anticipe que ses conditions d'utilisation ne seront vraisemblablement plus remplies pendant l'exécution de la manœuvre, le système de conduite automatisé doit : 1° Alerter le conducteur ; 2° Effectuer une demande de reprise en main ; 3° Engager et exécuter une manœuvre à risque minimal à défaut de reprise en main à l'issue de la période de transition ou en cas de défaillance grave.*

conduzione della vettura<sup>105</sup>. Siffatta impostazione è del tutto in linea con quanto prescritto dall'art. R416-6 c. route, ove si prevede che tutti i veicoli in movimento devono sempre avere un "conducteur". Così anche in caso di attivazione del sistema di guida autonoma è la stessa legge a sottolineare l'esigenza di una comunicazione tra uomo e macchina, per cui è l'uomo ad attivare il pilota automatico, il quale non può operare senza che chi è alla guida sia previamente informato del controllo che il sistema andrà ad assumere; la comunicazione resta costante per tutto il periodo di operatività del sistema, il quale deve informare il conducente in caso di malfunzionamento e richiedere a questi di riprendere il controllo della vettura.

È questa comunicazione uomo-macchina che fa transitare l'interprete dalla portata negativa dell'art. 123-1 c. route (cioè quella escludente la responsabilità penale) a quella positiva, che afferma la responsabilità penale del conducente. E infatti, la norma prescrive che laddove il conducente non risponda alla richiesta del sistema, con la quale gli viene ordinato di riprendere il controllo della vettura, tornerà ad applicarsi l'art. 121-1 c. route sulla responsabilità penale del conducente, che, quindi, opererà non solo quando il conducente ha il controllo dinamico del veicolo, ma anche quando avrebbe dovuto assumerlo e non lo ha fatto. In questo senso dispone il comma 3 dell'art. 123-1 del c. route, che ripristina il regime ordinario di responsabilità del conducente, confermando la necessità di una guida "inversamente assistita" e non sostituita dal sistema di IA<sup>106</sup>.

Ciò comporta che la responsabilità del conducente si presta a divenire una responsabilità omissiva allorché sia attivato il sistema di guida autonoma, poiché il guidatore assume una posizione di controllo di una fonte di pericolo e l'obbligo di attivarsi sorge chiaramente quando il sistema chieda al conducente di riprendere il controllo e questi non agisca in tal senso, come dispone chiaramente il comma 3 al n. 2.

---

<sup>105</sup> Lo scenario è inverso nel livello 2 di automazione, in cui la guida del conducente è assistita dal sistema

<sup>106</sup> Si riporta il testo del comma 3 dell'art. 123-1 c. route : « *Les dispositions du premier alinéa de l'article L. 121-1 sont à nouveau applicables :*

1° *Dès l'instant où le conducteur exerce le contrôle dynamique du véhicule à la suite d'une reprise en main de celui-ci ;*

2° *En l'absence de reprise en main du véhicule par le conducteur à l'issue de la période de transition faisant suite à une demande du système de conduite automatisé dans les conditions prévues au II de l'article L. 319-3 ;*

3° *Au conducteur qui ne respecte pas les sommations, injonctions ou indications données par les forces de l'ordre ou les règles de priorité de passage des véhicules d'intérêt général prioritaires prévues au présent code ».*

Ci si può allora chiedere se le forme di responsabilità omissiva vadano circoscritte a quella esplicitata nel testo (mancata ripresa di controllo della vettura a fronte di un'esplicita richiesta da parte del sistema), ovvero possa ritenersi sussistente anche quando nessuna richiesta pervenga dal sistema di guida autonoma. In altri termini: l'ordine che il sistema rivolge al conducente di riprendere il controllo della vettura limita la responsabilità omissiva al solo caso in cui il conducente non risponda a tale richiesta? Probabilmente no. La nozione di guida assistita disciplinata dalla normativa in esame riguarda il livello di autonomia 3, che denota un contesto di guida "eyes off", e non "mind off", come si desume chiaramente dal comma 2 dell'art. 123-1 c. route, che prevede che *"le conducteur doit se tenir constamment en état et en position de répondre à une demande de reprise en main du système de conduite automatisé"*. Ne deriva che al conducente è sempre richiesta una vigilanza quando il controllo dinamico è affidato al sistema di pilota automatico, per cui non si possono costruire delle aree di immunità penale del conducente per il solo fatto che a dirigere le operazioni di guida sia la tecnologia di IA. Questo dovere di vigilanza è necessario anche perché se, da un lato, si stima che la circolazione di auto a guida delegata riuscirà ad abbattere i livelli di pericolosità generati dalla guida umana attraverso una importante riduzione degli incidenti stradali, dall'altro, non è escluso che si assisterà a nuove forme di pericolosità connesse a situazioni improvvise che caratterizzano il caso concreto e per questo difficilmente prevedibili dal sistema. Non bisogna infatti sopravvalutare l'intelligenza artificiale: il prodotto tecnologico in questione è intelligente nel momento in cui conosce le regole sulla circolazione stradale e conforma il proprio agire alle stesse, ma lo è meno quando immagina che tutti gli utenti della strada tali regole rispetteranno. L'osservanza scrupolosa delle norme sulla circolazione stradale non è ipotizzabile da parte di tutti: se un siffatto atteggiamento conforme è immaginabile ad opera delle altre auto a guida autonoma, lo stesso non potrà dirsi per l'uomo, che spesso queste regole disattende. Ragionando con degli esempi, possiamo immaginare che il sistema di guida autonoma riuscirà a evitare collisioni con altri veicoli dovute a un mancato rispetto della distanza di sicurezza, in quanto sarà programmato per rispettare scrupolosamente il distanziamento imposto, tuttavia non potrà dirsi lo stesso nel caso di un soggetto che attraversi improvvisamente la strada e per giunta al di fuori delle strisce pedonali, tenendo così un comportamento non conforme alle regole sulla circolazione dei pedoni. L'evento di danno, difficilmente prevedibile dal pilota automatico, potrà però essere evitato dal guidatore, il quale magari riuscirà a prevedere quell'attraversamento (imprevedibile per la macchina)

poiché, ad esempio, sa di trovarsi in una zona in cui è presente un luogo di giochi per bambini (non segnalato e quindi non noto al sistema) e in cui a quell'ora è probabile che si stiano svolgendo attività ludiche.

I limiti della IA leggera, che non “ragiona” ma “calcola”, richiedono l'intervento di un “motore semantico”, che è l'uomo<sup>107</sup>, al quale potrà essere mosso un rimprovero per colpa per non aver agito pur disponendo di conoscenze causali superiori a quelle di un uomo medio<sup>108</sup>. In altri termini, mentre il livello di diligenza esigibile da parte del sistema di guida autonoma (e quindi da parte del produttore al momento dell'impostazione della conoscenza di esso) è quello tipico del conducente modello, che non dispone di conoscenze superiori a quelle mediamente richieste a ciascun conducente, diversa è la posizione dell'agente umano, il cui rimprovero può essere maggiormente individualizzato già al momento della valutazione dello standard di diligenza richiesto, esigendo una maggiore cautela in ragione delle maggiori conoscenze causali relative alla situazione concreta di cui eventualmente disponga.

Ritenuto allora che l'indicazione contenuta nell'art. 123-1, comma 3, n. 2 c. route sull'obbligo del conducente di riprendere il controllo della guida a fronte di una richiesta del sistema di pilota automatico non rende il primo esente dall'attuare una vigilanza costante anche quando una richiesta non pervenga dal sistema, potremmo ritenere che l'esplicitazione legislativa dell'obbligo di attivarsi assuma rilievo quantomeno sul piano della colpa, e non della responsabilità, la quale ultima sussisterebbe anche per omissioni diverse da quella legislativamente prevista. In altri termini, in questo caso la norma di rango legislativo pone una regola cautelare “precisa” (riprendere il controllo della guida se il sistema lo richiede) in grado di qualificare la *faute* come *délibérée*, che, come si è detto poco sopra, può essere contestata all'agente in presenza di due circostanze, cioè la violazione di una regola cautelare “precisa” prevista dalla legge o da un regolamento e la causazione indiretta dell'evento.

---

<sup>107</sup> L. Floridi, *La quarta rivoluzione. Come l'infosfera sta cambiando il mondo*, Milano, Raffaello Cortina Editore, 2017, secondo il quale “se le automobili che si guidano da sole potranno muoversi creando sempre meno problemi, (...) ciò non sarà dovuto al fatto che l'intelligenza artificiale forte sia stata infine realizzata, ma perché le “condizioni ambientali”, che automobili o droni devono affrontare, saranno rese sempre più adatte a un'intelligenza artificiale leggera e alle sue limitate capacità”, p. 166.

<sup>108</sup> Sulla rilevanza delle speciali conoscenze “causali” ai fini dell'innalzamento della misura oggettiva di diligenza, M. Romano, *Commentario sistematico del codice penale*, Milano, Giuffrè, 2004, *sub art. 43*, p. 459; G. Fiandaca, E. Musco, *Diritto penale, P.te gen.*, Bologna, Zanichelli, 2019, p. 585; G. Forti, *Colpa ed evento nel diritto penale*, Milano, Giuffrè, 1990, pp. 272 ss.

Ebbene, da questo potremmo dedurre che al conducente l'evento potrà essere attribuito a titolo di *faute délibérée* solo in caso di mancata ripresa del controllo del veicolo a fronte di una richiesta del sistema -sempreché il contributo causale apportato alla verifica dell'evento sia indiretto-, mentre laddove l'omissione si registri con riguardo a una regola cautelare diversa potrà essere contestata la *faute caractérisée* (che non richiede la violazione di una regola "precisa" prevista dalla legge o dal regolamento). Infine, ogni tipo di omissione (tanto di regola generica, quanto di regola precisa, sia essa prevista dalla legge o da una massima di esperienza) potrà essere contestata in caso di contributo causale diretto, per il quale è richiesta la *faute simple*.

## 5. La responsabilità del produttore

L'idea sottesa alla responsabilità del produttore è quella di sollecitare la produzione di sistemi con standard di sicurezza sempre più elevati<sup>109</sup> ed è con questa finalità che la riforma del 2021 prende in considerazione anche la posizione del produttore, disciplinandone la responsabilità nel nuovo art.123-2 c. route<sup>110</sup>. La norma completa il regime di responsabilità già tracciato dal precedente art. 123-1 c. route, prevedendo che il produttore o il suo rappresentante sono responsabili quando il veicolo è sottoposto al controllo dinamico del sistema di guida autonoma. Questo regime di responsabilità è però limitato a un catalogo chiuso di reati, riferendosi la norma solamente a tre *délits* del code pénal. Nella specie, il produttore è responsabile per il reato di omicidio involontario commesso dal conducente di un veicolo a motore con guida su strada (art. 221-6-1 c. pén.); per il reato di lesioni colpose all'integrità fisica da cui sia derivata un'incapacità lavorativa della durata superiore di tre mesi (art. 222-19-1 c. pén.), ovvero di durata pari o inferiore a tre mesi (art. 222-20-1 c. pén.).

---

<sup>109</sup> M. Monot-Fouletier, *Véhicule autonome: vers une autonomie du régime de responsabilité applicable ?*, Recueil Dalloz, 2018, p. 129.

<sup>110</sup> Art. 123-2 c. route : I. *Pendant les périodes où le système de conduite automatisé exerce le contrôle dynamique du véhicule conformément à ses conditions d'utilisation, le constructeur du véhicule ou son mandataire, au sens de l'article 3 du règlement (UE) 2018/858 du Parlement européen et du Conseil du 30 mai 2018, est pénalement responsable des délits d'atteinte involontaire à la vie ou à l'intégrité de la personne prévus aux articles 221-6-1, 222-19-1 et 222-20-1 du code pénal lorsqu'il est établi une faute, au sens de l'article 121-3 du même code.*  
II. *Sauf dans les cas prévus au 3° de l'article L. 123-1, lorsqu'une manœuvre effectuée par le système de conduite automatisé exerçant le contrôle dynamique du véhicule conformément à ses conditions d'utilisation contrevient à des règles dont le non-respect constitue une contravention, le constructeur du véhicule ou son mandataire, au sens de l'article 3 du règlement (UE) 2018/858 du Parlement européen et du Conseil du 30 mai 2018, est redevable pécuniairement de l'amende encourue.*

Ma vi è un altro requisito necessario affinché la responsabilità del produttore operi: occorre che sia accertata la colpa a mente dell'art. 121-3 c. pén., cioè la disposizione generale in tema di responsabilità, sopra analizzata, che pone le regole in tema di causalità e di *faute*.

Così predisposta, la norma sulla responsabilità del produttore presenta evidenti differenze rispetto alla norma sulla responsabilità del conducente. E infatti: i) il presupposto è che il controllo dinamico del veicolo sia in capo al sistema di guida autonoma; ii) la responsabilità penale è prevista limitatamente a tre disposizioni del codice penale; iii) è espressamente richiamata la norma sull'accertamento della colpa, che segue quindi i principi generali previsti dal code pénal.

i) La necessità che il sistema abbia il controllo dinamico del veicolo è un logico precipitato della delimitazione della responsabilità del conducente tracciata dall'articolo che precede quello in esame. Del resto, se il conducente non ha il controllo della vettura e un incidente si verifica, questo potrebbe essere conseguenza di un errore alla guida fatto dal sistema. La rinuncia all'individuazione di una responsabilità del sistema di IA è condivisa dal legislatore francese e, per questo, la responsabilità non poteva che ricadere sul produttore o sul suo rappresentante (*mandataire*). A mente del Regolamento (UE) 2018/858 è "rappresentante del costruttore" la persona fisica o giuridica stabilita nell'Unione che è nominata dal costruttore e che lo rappresenta dinanzi all'autorità di omologazione o all'autorità di vigilanza del mercato e agisce in suo nome<sup>111</sup>.

Il riferimento al rappresentante del costruttore è rilevante nel momento in cui lascia emergere l'attenzione del legislatore verso il destinatario di questa disposizione, che non potrebbe essere estesa al di là delle persone del produttore e del rappresentante, appunto. Restano fuori figure come il designer o l'addestratore, che comunque rivestono un ruolo fondamentale nel ciclo di produzione del sistema di IA. Sembra, allora, che questi soggetti saranno sì responsabili, ma secondo le norme ordinarie del codice penale e non alla luce delle disposizioni di nuova introduzione.

ii) Dopo aver delimitato l'ambito soggettivo di applicazione, la norma circoscrive il piano oggettivo, sancendo l'applicazione della relativa disciplina ai soli *délits* di omicidio

---

<sup>111</sup> art. 3, n. 41, Reg. (UE) 2018/858, relativo all'omologazione e alla vigilanza del mercato dei veicoli a motore e dei loro rimorchi, nonché dei sistemi, dei componenti e delle entità tecniche indipendenti destinati a tali veicoli.

e lesioni stradali colposi. Questa delimitazione del perimetro applicativo pone due questioni: la prima, sulla sorte della responsabilità in caso di reati diversi da quelli indicati; la seconda, sulla responsabilità dei soggetti diversi dal produttore e dal rappresentante.

Prendendo le mosse dal primo quesito, bisogna chiedersi, ad esempio, cosa accade se, quando il sistema ha il controllo dinamico, a seguito di uno scontro si danneggiano beni di terzi aventi natura patrimoniale e vi siano i presupposti per la contestazione del reato di danneggiamento (art. 322-1 c. pén). Il conducente potrebbe dimostrare che al momento dei fatti non aveva il controllo dinamico del veicolo e che quindi mancano i presupposti per dichiararne la responsabilità a mente dell'art. 123-1 c. de la route. Il difetto del sistema dovrebbe allora essere imputato al produttore, il quale però, a rigore, non dovrebbe rispondere per danneggiamento, ma solo in caso di omicidio o lesioni colposi. Così potrebbe sembrare che lo stesso legislatore abbia creato delle aree di irresponsabilità. Ma una simile conclusione è tuttavia affrettata. La norma si riferisce solo a tre reati che rientrano nei *délits*, mentre per gli altri, come ad esempio il reato di danneggiamento, deve applicarsi la logica del *code suiveur*, secondo la quale le aree non disciplinate dal code de la route tornano a essere sottoposte all'ordinaria disciplina del code pénal. Ne deriva che i *délits* diversi da quelli di omicidio e lesioni stradali e i crimes potranno essere imputati al produttore nei casi in cui il sistema esercitava un controllo dinamico se è provata la sua responsabilità secondo le regole generali del diritto penale, quindi nella specie se ne è provata la colpa. Quanto invece alla commissione di fattispecie contravvenzionali, come quella del superamento dei limiti di velocità consentiti (che, giova ancora ribadirlo, nel sistema francese è una sanzione penale e non amministrativa), una regola *ad hoc* è prevista dal c. 2 dello stesso articolo, che sarà analizzato a breve, in base alla quale il produttore sarà tenuto al pagamento dell'ammenda.

Arriviamo adesso alla seconda questione: la delimitazione oggettiva non vale per i soggetti diversi dal produttore e dal rappresentante, in quanto non sono destinatari della disciplina di nuova introduzione. Estendere nei loro confronti la responsabilità prevista dalla novella, nel silenzio della legge, potrebbe comportare un'interpretazione analogica *in malam partem* non tollerata in materia penale. Se non è d'aiuto l'estensione in via analogica della responsabilità del "*constructeur*", per individuare l'area di responsabilità dei suoi collaboratori che intervengono nella catena produttiva potremmo seguire una lettura in negativo dell'art. 123-2. Quest'ultimo percorso giunge tuttavia a conclusioni non condivisibili; infatti, la lettura in negativo comporterebbe di escludere l'accertamento della responsabilità di tali soggetti per i reati di lesioni e omicidio, ma si tratterebbe di

una forma di “immunità” penale difficile da giustificare: per quale ragione chi partecipa all’elaborazione dell’algoritmo dovrebbe sottrarsi a una tale responsabilità?

Piuttosto deve ammettersi che la responsabilità di soggetti diversi dal produttore e dal suo rappresentante può sussistere anche per i reati di omicidio e lesioni e deve essere accertata secondo le regole generali in materia penale, alla luce del meccanismo tipico del *code suiveur*, in base al quale ciò che non è previsto dal codice speciale si intende rimesso alla disciplina del *code pénal*.

A questo punto, se tanto per il produttore, quanto per soggetti da lui diversi vi è responsabilità per i reati di omicidio e lesioni stradali, per i quali occorre accertare che il sistema avesse il controllo della vettura e la presenza dei requisiti di cui all’art. 121-3 c. pén., bisogna chiedersi quale sia la portata innovativa dell’art. 123-2 c. de la route rispetto al regime ordinario di responsabilità penale. Apparentemente nessuna. Tuttavia possiamo provare a ragionare nell’ottica di una *ratio* che non è quella di escludere la responsabilità del produttore per i casi diversi da quelli indicati (cioè per reati diversi dall’omicidio e lesioni colpose stradali), ma di ancorare tale responsabilità per fatti gravissimi alla prova tanto del controllo dinamico del veicolo da parte del sistema, quanto della colpa del produttore. L’onere della prova per l’accusa è duplice: un requisito attinente al veicolo, un altro attinente alla colpa del produttore. Il primo è presupposto del secondo. Ed è vero che vi è tale rapporto di presupposizione tra controllo dinamico e colpa, infatti in assenza del nuovo art. 123-2 c. route il produttore sarebbe comunque responsabile se ne fosse provata la *faute*, che presuppone che il sistema di guida autonoma sia operativo e, quindi, che abbia il controllo dinamico del veicolo. Ma se il controllo dinamico è un presupposto che si sarebbe dovuto accertare anche senza un’apposita previsione in tal senso, allora la *ratio* della riforma può essere rinvenuta nella volontà del legislatore di ‘estrapolare’ tale requisito dal concetto generale di colpa, esplicitandolo. Così, la riforma potrebbe essere letta nel senso di evitare interpretazioni tendenti ad affermare la responsabilità del produttore per gravi reati quali l’omicidio colposo e le lesioni colpose anche quando non vi è un controllo totale del veicolo da parte del sistema.

Bisogna allora distinguere tra controllo totale e controllo parziale del veicolo da parte del sistema: per i reati di omicidio e lesioni solo il controllo totale fa scattare la responsabilità del produttore secondo la norma di nuova introduzione. Si tratta di una conclusione che sarebbe ragionevole estendere anche ai soggetti che contribuiscono all’ideazione dell’algoritmo, nel rispetto di un’interpretazione analogica *in bonam partem*.

Anche per costoro la responsabilità non può essere affermata se non quando il controllo dinamico è totalmente affidato all'intelligenza artificiale.

Laddove invece sia il conducente ad avere il controllo e questi si avvalga solo di un'assistenza da parte del sistema intelligente senza che a esso sia affidato il controllo dinamico totale della vettura, allora la responsabilità per i reati di omicidio e lesioni stradali non potrà essere fatta ricadere sul produttore.

iii) La colpa del produttore va accertata sulla base dell'art. 121-3 c. pén. (« *le constructeur du véhicule ou son mandataire (...) est pénalement responsable (...) lorsqu'il est établi une faute, au sens de l'article 121-3 du même code* »)

Il richiamo espresso manca nella norma sulla responsabilità del conducente, ma ciò non ne impedisce l'applicazione, poiché, come già ricordato, è la stessa natura di *code suiveur* che impone di leggere le norme del code de la route in combinato disposto con gli articoli contenuti nel code pénal, e a tale principio di certo non potrebbe fare eccezione la norma sulla *faute*, che ha una valenza generale per natura.

Il richiamo espresso contenuto all'art. 123-2 c. route (sulla responsabilità del produttore) non varrebbe, allora, a imporre una lettura dell'art. 123-1 c. route (sulla responsabilità del conducente) sganciata dall'art. 121-3 c. pén. (cioè della norma sulla *faute*) e quindi non potrebbe in alcun modo sostenersi che la norma sulla colpa valga solo là dove è richiamata, cioè solo per la responsabilità del produttore e non anche per la responsabilità del conducente.

Queste argomentazioni inducono a ritenere pleonastico il richiamo alla norma generale sull'accertamento della colpa, scelta di un legislatore premuroso, che forse vuole fugare ogni dubbio circa l'applicabilità del regime ordinario di responsabilità colposa. Ma, se così è, appare poco chiara la scelta di non prevedere un rinvio speculare in tema di responsabilità del conducente.

Probabilmente il richiamo espresso alle regole sulla responsabilità colposa si spiega perché quella del produttore è una figura che si presta all'introduzione di forme di responsabilità oggettiva occulta in quanto responsabilità da posizione. Il legislatore precisa invece che la responsabilità del produttore di un sistema intelligente presuppone sempre l'accertamento di un comportamento contrario a una regola cautelare, così da arginare ogni rischio di automatismo, che deriverebbe dallo stabilire una corrispondenza tra l'aver impostato il sistema di guida autonoma e la responsabilità del produttore, presumendo in questo modo la colpa di questi. A ritenere superfluo l'accertamento di una colpa nella produzione del manufatto intelligente, il produttore risponderebbe non più per

fatto proprio, bensì “per fatto altrui”. Una logica di questo tipo, che richiama la responsabilità da posizione, mal si concilierebbe con l’art. 121-1 c. pén, a mente del quale “*nul n’est responsable pénalement que de son propre fait*”<sup>112</sup>.

Il rischio di una responsabilità da posizione del produttore non è poi così remoto, basti pensare che la regola sulla responsabilità personale in passato ha conosciuto difficoltà applicative da parte della giurisprudenza proprio in tema di responsabilità dell’imprenditore, il quale veniva invece ritenuto responsabile “in modo automatico” dei fatti dei suoi preposti in quanto garante della corretta applicazione delle norme di legge in ambito aziendale<sup>113</sup>. Ciò comportava una responsabilità che, sebbene non oggettiva, era comunque presunta. Per questo motivo è intervenuto il legislatore nel 1976 che ha modificato il precedente art. L. 263-2 code du travail, prevedendo che il reato è imputabile all’imprenditore se ha violato le regole sull’igiene e sulla sicurezza “*par faute personnelle*”. In questo modo il legislatore ha precisato che occorre una responsabilità personale, cioè per fatto proprio dell’imprenditore<sup>114</sup>, e la specificazione non è di poco momento in un sistema in cui, “secondo la dottrina prevalente, il principio di colpevolezza, inteso come necessità di un elemento soggettivo, non riceve un riconoscimento costituzionale”<sup>115</sup>.

Questa evoluzione storica che ha interessato la responsabilità dell’imprenditore potrebbe allora spiegare la scelta del rinvio espresso alla norma sull’accertamento dell’elemento soggettivo per la responsabilità del produttore del sistema di guida autonoma, per il quale il rischio è quello di innescare una presunzione di colpa in presenza del controllo del veicolo da parte del sistema di IA.

Grazie al rinvio all’art. 121-3 c. pén. risulta certa l’applicazione della stessa norma e, in particolare, del suo comma 4, in tema di contribuito causale indiretto, posto che in

---

<sup>112</sup> La norma positivizza il principio di personalità della responsabilità penale (*principe de personnalité*), sull’argomento si rinvia in generale a X. Pin, *Droit pénal général*, 12 ed., Paris, Dalloz, 2020, p. 28.

<sup>113</sup> E. Dreyer, *Droit pénal*, cit., p. 818.

<sup>114</sup> La riforma interviene in risposta all’orientamento giurisprudenziale che riconosceva in via automatica la responsabilità dell’imprenditore. Ciononostante, anche dopo la riforma l’atteggiamento della giurisprudenza si è mosso verso un’interpretazione restrittiva della norma, per cui la responsabilità dell’imprenditore era presunta, con conseguente superamento della presunzione solo attraverso prova contraria. La dottrina ha evidenziato che tale orientamento giurisprudenziale era finalizzato a colmare il vuoto dovuto all’assenza di una legge sulla responsabilità penale delle persone giuridiche. Dopo l’introduzione di questa forma di responsabilità, il datore di lavoro coincide con l’ente e a questi spetta la vigilanza sul rispetto delle norme in materia di sicurezza sul lavoro. L’ente è responsabile se il reato è commesso dai suoi organi o rappresentati nel suo interesse (*pour son compte*). Secondo questa opinione, la conseguenza è che oggi per la responsabilità dell’imprenditore non si può più prescindere dalla prova della sua colpa da parte dell’accusa (E. Dreyer, *op. cit.*, p. 823).

<sup>115</sup> R. Sicurella, *cit.*, p. 974.

questa manifestazione si porrà la condotta del produttore rispetto all'evento morte o lesioni. Di conseguenza, la colpa da accertarsi è la c.d. *faute qualifiée*, nelle due declinazioni della *faute délibérée* e della *faute caractérisée*.

Su questo punto va sottolineato che i *délits* di omicidio e lesioni stradali prevedono un aggravamento di pena ove sia accertata la *faute délibérée*, la quale, come si è poco sopra detto, sussiste in quanto sia violata una regola di condotta precisa nei suoi contenuti e prevista da una legge o da un regolamento. In assenza di una normativa specifica che disciplini la produzione di sistemi di guida autonoma, forse è più probabile che sia contestata la *faute caractérisée*.

### 5.1 Il regime delle contravvenzioni. Responsabilità per il fatto del terzo?

È con un atteggiamento duplice che la dottrina e la giurisprudenza francesi cercano di domare le tensioni suscitate nell'ordimento francese dall'assenza di un parametro costituzionale che sancisca il principio di colpevolezza: da un lato, si invoca il principio della presunzione di innocenza, di cui all'art. 9 della Dichiarazione dei diritti dell'uomo e del cittadino; dall'altro, si valorizzano proprio gli artt. 121-3 c. pén. sull'elemento soggettivo del reato, e 121-1 c. pén sul *principe de personnalité*.

Nondimeno, forme di responsabilità oggettiva 'occulta' permangono e se ne ha riprova già dalla lettura dell'art. 121-3, ultimo comma, c. pén in materia di contravvenzioni, per le quali responsabilità è esclusa "*en cas de force majeure*", mentre nulla si dice con riguardo all'elemento soggettivo, cosicché dubbi sorgono sul carattere personale e soggettivo della responsabilità per questa categoria di reati<sup>116</sup>.

L'interesse verso tali questioni emerge in questa sede se si considera che forme di *responsabilité pénale du fait d'autrui* sono rintracciabili nel codice della strada francese, che qualifica come contravvenzionali (e quindi penali) infrazioni che nel nostro ordinamento hanno natura amministrativa. Orbene, è vero che, intervenendo in materia penale, anche il code de la route prevede come principio generale quello della responsabilità personale del conducente (art. 121-1 c. route: "*le conducteur d'un véhicule est responsable pénalement des infractions commises par lui dans la conduite dudit*

---

<sup>116</sup> La ricostruzione delle opinioni dottrinali circa la *faute contraventionnelle* è affrontata nella dottrina italiana da R. Sicurella, *cit.*, pp. 975 ss.

*véhicule*”), ma tale principio conosce due deroghe<sup>117</sup>, alle quali da ultimo se ne è aggiunta una terza proprio per effetto della recente riforma.

La prima deroga è prevista dall’art. 121-2 c. route, sulla “responsabilità pecuniaria” del titolare del certificato di immatricolazione del veicolo con riferimento alle contravvenzioni indicate dalla stessa disposizione, tra cui la contravvenzione della sosta in zona vietata o della sosta oltre il limite temporale consentito<sup>118</sup>. La dottrina ha subito messo in evidenza la *ratio* sottesa all’introduzione della norma: semplificare la prova della pubblica accusa con riferimento a contravvenzioni accertate pressoché sempre in assenza del conducente del veicolo<sup>119</sup>. Ma la responsabilità non resta ferma in ogni caso, poiché « *le titulaire du certificat d'immatriculation du véhicule est responsable pécuniairement (...) à moins qu'il n'établisse l'existence d'un événement de force majeure ou qu'il ne fournisse des renseignements permettant d'identifier l'auteur véritable de l'infraction* ». Quindi è ammessa la prova liberatoria relativa al verificarsi di cause di forza maggiore o, in alternativa, il titolare del certificato di immatricolazione deve addurre degli elementi che consentano l’individuazione dell’effettivo responsabile.

La seconda deroga al principio di responsabilità personale è prevista dal successivo art. 121-3 c. route<sup>120</sup>, a mente del quale: « *Par dérogation aux dispositions de l'article L. 121-1, le titulaire du certificat d'immatriculation du véhicule est redevable pécuniairement de l'amende encourue pour des infractions dont la liste est fixée par décret en Conseil d'Etat, à moins qu'il n'établisse l'existence d'un vol ou de tout autre événement de force majeure ou qu'il n'apporte tous éléments permettant d'établir qu'il n'est pas l'auteur véritable de l'infraction* ». Nonostante la stessa *ratio* ispiratrice (posto che anche qui il legislatore persegue la finalità di semplificare l’accertamento della

---

<sup>117</sup> L. Desessard, *op. cit.*, p. 21 (con riferimento all’art. 121-2) e p. 25 (per l’art. 121-3).

<sup>118</sup> Art. 121-2 : I. *Par dérogation aux dispositions de l'article L. 121-1, le titulaire du certificat d'immatriculation du véhicule est responsable pécuniairement des infractions à la réglementation sur le stationnement des véhicules ou sur l'acquittement des péages pour lesquelles seule une peine d'amende est encourue ainsi que des contraventions relatives à l'abandon d'ordures, de déchets, de matériaux ou d'autres objets, à moins qu'il n'établisse l'existence d'un événement de force majeure ou qu'il ne fournisse des renseignements permettant d'identifier l'auteur véritable de l'infraction.*

<sup>119</sup> L. Desessard, *op. cit.*, p. 21.

<sup>120</sup> Per un commento sulla norma, tra gli altri, J.P. Céré, *L'illusion pénale de la responsabilité du titulaire du certificat d'immatriculation*, AJ pén., sept. 2019, pp. 418-421 ; L. Desessard, *Le traitement simplifié des infractions à la circulation routière*, in *Un droit pénal postmoderne ?*, Paris, Presses Universitaires de France, pp. 298 ss.

responsabilità al fine di non lasciare impunte certe condotte stante la difficoltà di identificarne l'autore)<sup>121</sup>, a differenza che nella disposizione precedente, la scelta di delegare la delimitazione dell'ambito di applicazione della norma a una previsione secondaria consente di estenderne la portata sulla base di un procedimento più agevole rispetto a quello legislativo. In particolare, il catalogo dei reati cui si riferisce la norma è indicato all'art. R.121-6 c. route<sup>122</sup>, che fa riferimento a delle contravvenzioni e a un unico delitto, quello di circolazione in assenza di copertura assicurativa<sup>123</sup>.

Anche qui però è ammessa la prova contraria: il titolare del certificato di immatricolazione può liberarsi se prova un avvenuto furto (*l'existence d'un vol*) o altro evento di forza maggiore (*autre événement de force majeure*) o comunque se dimostra che lui non è l'autore dell'infrazione (*éléments permettant d'établir qu'il n'est pas l'auteur véritable de l'infraction*).

---

<sup>121</sup> G. Defrance, *Le propriétaire du véhicule redevable pénalement*, *Jur. Automobile*, n. 745, oct. 2003, p. 485.

<sup>122</sup> Art. R121-6 : *Le titulaire du certificat d'immatriculation du véhicule est, en application de l'article L. 121-3, redevable pécuniairement de l'amende encourue pour des infractions aux règles sur :*

1° *Le port d'une ceinture de sécurité homologuée dès lors que le siège qu'il occupe en est équipé prévu à l'article R. 412-1 ;*

2° *L'usage du téléphone tenu en main ou le port à l'oreille de tout dispositif susceptible d'émettre du son prévus à l'article R. 412-6-1 ;*

3° *L'usage de voies et chaussées réservées à certaines catégories de véhicules, de voies vertes et d'aires piétonnes prévu au II de l'article R. 412-7 ;*

4° *L'arrêt, le stationnement ou la circulation sur les bandes d'arrêt d'urgence prévus à l'article R. 412-8, au 9° du II de l'article R. 417-10 et à l'article R. 421-7 ;*

5° *Le respect des distances de sécurité entre les véhicules prévu à l'article R. 412-12 ;*

6° *Le franchissement et le chevauchement des lignes continues prévus aux articles R. 412-19 et R. 412-22 ;*

6° bis *Le sens de la circulation ou les manœuvres interdites prévus aux articles R. 412-28 et R. 421-6 ;*

7° *Les signalisations imposant l'arrêt des véhicules prévues aux articles R. 412-30, R. 412-31 et R. 415-6 ;*

8° *Les vitesses maximales autorisées prévues aux articles R. 413-14, R. 413-14-1 et R. 413-17 ;*

9° *Le dépassement prévu aux articles R. 414-4, R. 414-6 et R. 414-16 ;*

10° *L'engagement dans une intersection ou dans l'espace compris entre les deux lignes d'arrêt prévu à l'article R. 415-2 ;*

10° bis *La priorité de passage à l'égard du piéton prévue à l'article R. 415-11 ;*

11° *L'obligation du port d'un casque homologué d'une motocyclette, d'un tricycle à moteur, d'un quadricycle à moteur ou d'un cyclomoteur prévue à l'article R. 431-1 ;*

12° *L'obligation, pour faire circuler un véhicule terrestre à moteur, d'être couvert par une assurance garantissant la responsabilité civile, prévue aux articles L. 211-1 et L. 211-2 du code des assurances et à l'article L. 324-2 ;*

13° *Le port de plaques d'immatriculation dans les conditions prévues à l'article R. 317-8.*

<sup>123</sup> Si tratta di un delitto in quanto punito con un'ammenda di € 3.750 che è il valore minimo dell'ammenda fissata per i delitti ai quali può essere applicata anche la sanzione dell'*emprisonnement* per un massimo di 10 anni, da sola o congiunta alla pena pecuniaria, cfr. E. Dreyer, *Droit pénal*, cit., p. 182.

Le due norme istituiscono delle presunzioni di responsabilità<sup>124</sup> e per questo si è detto che apportano una deroga al principio della responsabilità personale del conducente. Esse presentano delle differenze non irrilevanti: mentre l'art. 121-2 c. route prevede che il proprietario del veicolo « *est responsable pécuniairement des infractions pour lesquelles une peine d'amende est encourue* », l'art. 121-3 c. route afferma che il titolare del certificato di immatricolazione è “obbligato” al pagamento dell'ammenda (“*redevable pécuniairement*”). Più avanti la norma ha altresì cura di precisare che questi *non è penalmente responsabile del reato* e infatti, la sentenza del tribunale che condanna al pagamento dell'ammenda non è iscritta nel casellario giudiziale, né tantomeno potrà tale condanna essere valutata ai fini di un eventuale riconoscimento della recidiva<sup>125</sup>. È stato allora detto che non si tratta di una responsabilità penale in senso stretto, ma di una responsabilità pecuniaria<sup>126</sup>, mentre altri hanno notato che questa precisazione terminologica sarebbe stata voluta dal legislatore per evitare tensioni col principio della responsabilità penale personale<sup>127</sup>.

Come anticipato, una nuova disposizione simile a queste di cui si è appena detto – che come esse sembra andare in una direzione opposta al principio della responsabilità personale- è stata introdotta dalla riforma del 2021. Si tratta del nuovo art. L123-2 c. route, comma secondo, che prevede: « *Sauf dans les cas prévus au 3° de l'article L. 123-1, lorsqu'une manœuvre effectuée par le système de conduite automatisé exerçant le contrôle dynamique du véhicule conformément à ses conditions d'utilisation contrevient à des règles dont le non-respect constitue une contravention, le constructeur du véhicule ou son mandataire, au sens de l'article 3 du règlement (UE) 2018/858 du Parlement européen et du Conseil du 30 mai 2018, est redevable pécuniairement de l'amende encourue* ».

---

<sup>124</sup> F. Desportes, F. Le Guehec, *Droit pénal général*, cit., p. 481 ; Y. Mayaud, *Droit pénal général*, 6ed., Paris, Presses Universitaires de France, 2018, pp. 448-450.

<sup>125</sup> Si tratta del comma II, che così statuisce: « *La personne déclarée redevable en application des dispositions du présent article n'est pas responsable pénalement de l'infraction. Lorsque le tribunal, y compris par ordonnance pénale, fait application des dispositions du présent article, sa décision ne donne pas lieu à inscription au casier judiciaire, ne peut être prise en compte pour la récidive et n'entraîne pas retrait des points affectés au permis de conduire. Les règles sur la contrainte judiciaire ne sont pas applicables au paiement de l'amende* ».

<sup>126</sup> G. Defrance, *op. cit.*, p. 486.

<sup>127</sup> L. Desessard, *Le traitement simplifié des infractions*, *op. cit.*, p. 298.

La norma riportata afferma l'obbligo del pagamento dell'ammenda da parte del produttore o del rappresentante quando il sistema che ha il controllo dinamico del veicolo viola delle norme che danno luogo a una contravvenzione. Il produttore risponde, ma di un'azione che non è propria, poiché non è lui ad aver agito in modo contrario alla legge, bensì il sistema.

La norma non fa però riferimento alla responsabilità penale, tanto è vero che afferma che il produttore è *redevable pécuniairement*.

La terminologia impiegata è totalmente sovrapponibile all'art. 121-3 c. route di cui si è detto sopra, in quanto non si fa riferimento alla "responsabilità", bensì alla *redevabilité*. Tuttavia, la norma di nuova introduzione omette qualsiasi richiamo espresso all'art. 121-3 c. route, con la conseguenza che allo stato non risultano applicabili le cause di esonero dell'obbligo del pagamento dell'ammenda che formano oggetto della prova liberatoria in capo al titolare del certificato di immatricolazione, salvo sostenerne un'estensione in via analogica. Ciò significa che il produttore non potrà addurre a sua difesa il furto del veicolo o gli eventi di forza maggiore di cui l'art. 121-3 c. route tiene conto, né potrà liberarsi provando che lui non è l'autore dell'infrazione. Ma, del resto, la soluzione non è incoerente, poiché qui obbligato è il produttore, soggetto cui sarebbe del tutto irragionevole estendere la prova liberatoria di cui all'art. 121-3 c. route, che, fatto salvo il richiamo alle generiche cause di forza maggiore, è costruita attorno alla figura del titolare del certificato di immatricolazione. Così, se ha senso ritenere questi liberato dall'obbligo di pagare la sanzione pecuniaria in caso di contravvenzione che sia posta in essere da chi si è illegittimamente impossessato della vettura, lo stesso non varrebbe per il produttore, il quale è chiamato in gioco sia che alla guida vi sia il legittimo proprietario della vettura, sia allorché a condurre quest'ultima sia l'autore del furto della stessa. Allo stesso modo, sarebbe del tutto irragionevole ritenere il produttore liberato laddove riesca a provare che lui non è l'autore dell'infrazione, poiché questo è un dato che si dà già per assodato in quanto è certo che non era lui alla guida del veicolo.

La logica sottesa alla nuova disposizione è diversa da quella che ha ispirato queste due cause di esonero della "responsabilità" del titolare del certificato di immatricolazione, e coincide, piuttosto, con la scelta di addossare al produttore i costi di un'infrazione realizzatasi a causa di un difetto del sistema. Tanto è vero che la norma precisa che l'infrazione deve verificarsi quando l'impiego del sistema di guida autonoma avviene "*conformément à ses conditions d'utilisation*" e, ciò malgrado, la violazione della normativa si verifichi.

Scartate la prova di un avvenuto furto e quella sull'estraneità nella realizzazione dell'infrazione, non resta che la sopravvenienza di una causa di forza maggiore, unica condizione di esclusione della *redevabilité* prevista per la figura del titolare del certificato di immatricolazione che è possibile applicare in via analogica al costruttore<sup>128</sup>.

Tuttavia, nonostante sia possibile questa limitata applicazione analogica delle cause di esonero della *redevabilité* di cui all'art. 121-3 c. route, l'assenza di qualsivoglia prova liberatoria nella disposizione di nuova introduzione (art. 123-2 c. route) potrebbe porre il problema della sua compatibilità col principio del carattere personale della responsabilità penale, come tenterò di spiegare di seguito.

## **5.2 Il dibattito sorto intorno all'art. 121-3 del code de la route e la compatibilità con i principi della responsabilità personale e della presunzione di innocenza**

Di deroga al principio della responsabilità personale per fatto proprio del conducente la dottrina francese aveva già iniziato a discutere ancor prima che entrasse in vigore il nuovo code de la route. Difatti, la disciplina oggi contenuta agli artt. 121-2 e 121-3 c. route riprende, rispettivamente, i vecchi artt. L21-1 e L21-2 del precedente codice; proprio all'indomani dell'introduzione di quest'ultima norma (art. L21-2), per effetto della legge n. 99-505 del 18 giugno 1999, le perplessità furono tali da sfociare in una questione di legittimità costituzionale deferita al Conseil constitutionnel prima della promulgazione della legge stessa<sup>129</sup>. Ciò che si censurava era proprio la tensione con i

---

<sup>128</sup> Un'altra soluzione potrebbe essere quella di applicare analogicamente le cause di esonero di responsabilità di cui all'art. 121-2 c. route dedicato alla responsabilità pecuniaria del titolare del certificato di immatricolazione. Tale norma (v. *supra* in questo paragrafo) fa dipendere l'esonero della responsabilità dalla prova di una causa di forza maggiore ovvero dalla collaborazione mostrata all'autorità giudiziaria che porti all'individuazione dell'autore dell'infrazione ed è applicabile anche agli enti e non solo alle persone fisiche, quindi potrebbe estendersi all'impresa produttrice. Tuttavia, non mancano argomenti per ritenere che tale conclusione non sia percorribile, primo fra tutti il dato letterale, posto che mentre l'art. 121-2 c. route si riferisce alla responsabilità pecuniaria, la norma di nuova introduzione (art. 123-2 c. route) definisce il produttore *redevable pécuniairement*.

<sup>129</sup> Il deferimento *ex ante*, cioè prima della promulgazione della legge, è previsto dall'art. 61 Cost. Nel caso di specie la decision n. 99-411 del 16 giugno 1999 del Cons. cost. ([www.conseil-constitutionnel.fr](http://www.conseil-constitutionnel.fr)) era stata sollecitata da un gruppo di sessanta deputati. Con essa il Conseil Constitutionnel ha dichiarato conforme alla Costituzione l'art. 6 della legge n. 99-505 del 18 giugno 1999, che, a modifica dell'allora vigente c. route, introduceva l'art. 21-2 che prevedeva che il titolare del certificato di immatricolazione fosse tenuto al pagamento dell'ammenda per talune contravvenzioni indicate dalla stessa norma (tra cui l'eccesso di velocità), salva la prova di un avvenuto furto, di una causa di forza maggiore (si riporta di seguito il testo : « I - Par dérogation aux dispositions de l'article L. 21, le titulaire du certificat d'immatriculation du véhicule est redevable pécuniairement de l'amende encourue pour des contraventions à la réglementation sur les vitesses maximales autorisées et sur les signalisations imposant l'arrêt des véhicules, à moins qu'il n'établisse l'existence d'un vol ou de tout autre événement de force majeure ou qu'il n'apporte tous éléments permettant d'établir qu'il n'est pas l'auteur véritable de l'infraction. II - La personne déclarée redevable en

principi della responsabilità penale personale e della presunzione di innocenza, posto che la norma introduceva una presunzione di responsabilità<sup>130</sup>.

Il contrasto col principio della responsabilità penale personale è stato escluso dal Conseil constitutionnel, che ha dichiarato la disposizione allora vigente conforme all'art. 9 della Dichiarazione del 1789, che istituisce la presunzione di innocenza in materia penale<sup>131</sup>. La decisione conserva la sua attualità ancora oggi poiché la norma è stata riportata nell'art. 121-3 c. route del 2001, pertanto le conclusioni raggiunte dal Conseil constitutionnel sono ora estensibili alla nuova disposizione.

Per sostenere la conformità di questa forma di responsabilità con i principi della Costituzione, il ragionamento del Conseil Constitutionnel fa leva su tre punti. In primo luogo, il Consiglio osserva che, sebbene l'art. 9 della Dichiarazione dei diritti dell'uomo e del cittadino impedisca al legislatore di istituire la presunzione di colpa, in materia contravvenzionale tali presunzioni sono ammesse, seppur in via eccezionale. Il Conseil precisa le caratteristiche che tali presunzioni devono avere, affermando che: a) deve trattarsi di presunzioni semplici, cioè vincibili attraverso la prova contraria; b) tali presunzioni devono essere riferite a fatti che, così come verificatisi, rendono verosimile

---

*application des dispositions du présent article n'est pas responsable pénalement de l'infraction. Lorsque le tribunal de police, y compris par ordonnance pénale, fait application des dispositions du présent article, sa décision ne donne pas lieu à inscription au casier judiciaire, ne peut être prise en compte pour la récidive et n'entraîne pas retrait des points affectés au permis de conduire. Les règles sur la contrainte par corps ne sont pas applicables au paiement de l'amende. III- Les deuxième et troisième alinéas de l'article L. 21-1 sont applicables dans les mêmes circonstances » ). Per un commento alla decisione, oltre al relativo *Commentaire de la décision n. 99-411 DC du 16 juin 1999, Les Cahiers du Conseil constitutionnel, ch. 7* ([www.conseil-constitutionnel.fr](http://www.conseil-constitutionnel.fr)), si vedano B. Mathieu, *La sécurité routière fait progresser le droit constitutionnel pénal*, *Recueil Dalloz*, 15 juill 1999, n. 27 ; F. Luchaire, *Deux décisions du Conseil constitutionnel appelées à faire jurisprudence ?*, *Rev. Dr. Publ.*, 5, 1999, p. 1283-1292 ; J. Buisson, *op. cit.* ; Y. Mayaud, *Entre le dit et le non-dit, ou les leçons de droit pénal du Conseil constitutionnel*, *Recueil Dalloz*, 1999, p. 589.*

<sup>130</sup> Non si tratta di un'eccezione. Il codice penale francese ancor oggi contiene dei reati basati su presunzioni di responsabilità. È questo il caso dell'art. 225-6 c. pén. in materia di *proxénétisme* (delitto assimilabile allo sfruttamento della prostituzione nell'ordinamento italiano), che considera accertato l'elemento materiale del reato a carico del soggetto che, vivendo o intrattenendo una relazione abituale con persone dedite alla prostituzione, non sia in grado di giustificare il proprio tenore di vita; ancora l'art. 321-6 in materia di ricettazione (*recel*), che istituisce una presunzione a carico di chi, intrattenendo una relazione abituale con una o più persone dedite alla commissione di reati da cui derivi un profitto diretto o indiretto, non sia in grado di giustificare il proprio tenore di vita (per questo ultimo reato Cass. crim., 25 jan. 2012, n. 11-90.117 ha ritenuto di non sollevare la questione di costituzionalità). Per altri esempi di presunzioni di diritto previste dalla legislazione speciale, si rinvia a J. Buisson, *Les Présomptions de culpabilité, Procédures – Ed. Juris-Classeur*, 12, 1999, pp. 3-4.

<sup>131</sup> Art. 9. *Tout homme étant présumé innocent jusqu'à ce qu'il ait été déclaré coupable, s'il est jugé indispensable de l'arrêter, toute rigueur qui ne serait pas nécessaire pour s'assurer de sa personne doit être sévèrement réprimée par la loi.*

la presunzione di imputabilità (posto che in generale il conducente è anche proprietario del veicolo<sup>132</sup>); c) in ogni caso deve essere assicurato l'esercizio del diritto di difesa<sup>133</sup>.

In secondo luogo, precisa il Consiglio che la condanna postula una “*faute personnelle*” poiché presuppone che non sia raggiunta la prova contraria (e in questo caso la *faute personnelle* consiste nel difetto di vigilanza da parte dell'imputato), ovvero che l'imputato si sia rifiutato di ammettere la propria responsabilità (ipotesi in cui la *faute personnelle* consiste nel rifiuto a contribuire a ricostruire la verità)<sup>134</sup>.

In terzo luogo e da ultimo, il Consiglio fa leva sul dato letterale, che esclude espressamente tanto la natura penale di tale responsabilità, quanto l'iscrizione nel casellario giudiziale e afferma l'irrelevanza della condanna ai fini della recidiva.

Secondo una parte della dottrina, la decisione del Conseil constitutionnel ammette forme di “*présomption de culpabilité*”, ma entro limiti assai ristretti e nel rispetto del diritto di difesa in giudizio<sup>135</sup>. Il Consiglio costituzionale, infatti, fissa tre presupposti (*supra*, lett. *a*, *b*, *c*) affinché la presunzione di colpevolezza non si scontri con la presunzione di innocenza e, in caso di responsabilità, richiede comunque una “*faute*”, cioè la colpa del soggetto<sup>136</sup>. Ma ciò che è importante osservare è che tale ragionamento è ristretto alla categoria delle contravvenzioni, le sole infrazioni per le quali, a mente dell'art. 9 della Dichiarazione del 1769, è possibile derogare alla presunzione di innocenza.

Anche la giurisprudenza successiva alla sentenza del 1999 è solita riferirsi a tale presunzione definendola come *présomption de culpabilité*, che in materia di contravvenzioni è ammessa, purché non sia assoluta e l'applicazione della pena non sia quindi frutto di un automatismo<sup>137</sup>.

---

<sup>132</sup> Come sottolinea G. DeFrance, *op. cit.*, p. 487.

<sup>133</sup> Sottolinea tuttavia J. Buisson che il contenuto della decisione lascia presumere che il Conseil constitutionnel non intende limitare tali presunzioni alla sola materia contravvenzionale. *Op. cit.*, p. 3.

<sup>134</sup> Par. 7 della sentenza.

<sup>135</sup> F. Luchaire, *op. cit.*, p. 1283 e p. 1287 ss.

<sup>136</sup> Certamente si tratta di una pronuncia che non è andata esente da critiche. Anzi, la dottrina ha avuto modo di evidenziare che siffatta soluzione dà luogo a contraddizioni poiché il proprietario sarà responsabile se, conoscendo l'identità dell'autore della contravvenzione non lo denuncia; tuttavia, anche laddove sia in possesso di tale informazione, non potrà essere punito se riesce a dimostrare che non è lui l'autore dell'infrazione dichiarando e provando che si trovava in viaggio in quel momento in un altro continente. F. Luchaire, *op. cit.*, p. 1289.

<sup>137</sup> Cass. Crim. 5 janv. 2011, n. 10-90.112, pronunciata con riferimento all'art. L121-3 c. route introdotto dalla loi n. 2009-526 del 12 maggio 2009, affermava « [...] *la disposition légale critiquée ne porte pas atteinte au principe de la présomption d'innocence, dès lors que des présomptions de culpabilité peuvent*

Una dottrina più attenta offre invece una lettura probabilmente più esatta della disposizione, partendo dal presupposto che nelle contravvenzioni l'*élément moral* è sempre presunto. Così, si è precisato che la presunzione introdotta dall'allora art. 21-2 (oggi art. 121-3 c. route) non attiene alla colpevolezza, ma all'elemento materiale dell'infrazione (cioè al contributo causale del titolare del certificato di immatricolazione)<sup>138</sup>, atteso che le contravvenzioni già si caratterizzano per la debolezza della componente psicologica che, in assenza di disposizione contraria, è presunta; per ogni contravvenzione, difatti, è sufficiente dimostrare che il fatto è materialmente imputabile all'agente<sup>139</sup> e non è richiesta la prova dell'elemento soggettivo, salvo che la legge lo richieda<sup>140</sup>. E allora, non essendo possibile presumere ciò che di regola è già presunto (cioè la *culpabilité*), la presunzione introdotta dal legislatore va riferita a ciò che di regola va invece provato (cioè il contributo causale del proprietario), col risultato che l'art. 121-3 c. route introdurrebbe una *presunzione di responsabilità*, mentre normalmente per le contravvenzioni opera solo una *presunzione di colpevolezza*. Si tratta di una lettura più convincente, che dà un senso alla presunzione legislativa, la quale altrimenti sarebbe superflua perché non farebbe altro che ribadire quanto è già *acquis* in materia di contravvenzioni.

Oltre all'oggetto su cui ricade la presunzione (*présomption de culpabilité* ovvero de *responsabilité*), si discute anche della natura della responsabilità del proprietario della vettura.

Il dibattito è sicuramente alimentato dalla lettera della norma, la quale afferma espressamente che il titolare del certificato di immatricolazione non è penalmente

---

*être établies, notamment en matière contraventionnelle, lorsqu'elles ne revêtent pas de caractère irréfragable et qu'est assuré le respect des droits de la défense ».*

<sup>138</sup> È questa l'opinione di Y. Mayaud, *op. cit.*, il quale precisa che: « *le législateur ne peut pas innover sur ce qui est acquis* », facendo cioè riferimento alla presunzione di *culpabilité* già ammessa in materia contravvenzionale.

<sup>139</sup> Y. Mayaud, *Droit pénal général, cit.*, p. 352; E. Dreyer, *Droit pénal, cit.*, p. 81. Ma è anche vero che questa non è una caratteristica propria di tutte le contravvenzioni. Infatti, bisogna isolare da questa categoria le contravvenzioni che prima erano classificate come *délits* e che si trovano principalmente nella parte regolamentare del code pénal e contemplano la componente morale dell'infrazione, che va provata dalla pubblica accusa. Allora, la presunzione di colpevolezza va ristretta alle sole contravvenzioni che rappresentano mere irregolarità, prive di un elevato disvalore sociale (E. Dreyer, *op. cit.*, pp. 184 e 185); G. Defrance, *op. cit., Jur. Aut. 2003*, p. 486.

<sup>140</sup> Come avviene ad esempio per la contravvenzione di cui all'art. R. 625 c. pén. sulle lesioni che comportano una incapacità lavorativa inferiore o pari a otto giorni, che sono "*violences volontaires*", F. Desportes, F. Le Guehec, *op. cit.*, p. 428.

responsabile. Sul punto si è osservato che la norma contiene in sé un'incoerenza teorica<sup>141</sup> poiché prima pone a carico del proprietario della vettura la contravvenzione, che ha natura penale, e subito dopo esclude il carattere penale di tale responsabilità.

Pur a fronte di un dato letterale così chiaro, la dottrina maggioritaria riconosce a questa responsabilità natura penale<sup>142</sup>, anche alla luce del carattere punitivo o comunque quasi penale della sanzione<sup>143</sup>. Del resto, non possono che accogliersi gli argomenti di carattere sostanziale posti a sostegno di siffatta opinione, che fanno principalmente leva, oltre che sul tipo di sanzione, cioè l'ammenda, che postula la commissione di un reato, anche sul presupposto per l'applicazione di tale sanzione, che è l'aver commesso una contravvenzione, quindi un reato<sup>144</sup>.

Dello stesso avviso era il Governo, che nelle osservazioni presentate in occasione della questione rimessa al Conseil constitutionnel, pur non affermando espressamente la natura penale di tale forma di punibilità, aveva comunque evidenziato il carattere punitivo e repressivo della sanzione, che imponeva il rispetto dei principi propri della materia penale, tra i quali la presunzione di innocenza<sup>145</sup>. E, ancora, anche dalla sentenza del Conseil constitutionnel sembra emergere, seppur solo implicitamente, la natura penale di tale responsabilità; d'altra parte, il problema della costituzionalità di forme di responsabilità presunta non può che porsi là dove opera il principio di presunzione di innocenza, vale a dire entro il perimetro delle sanzioni penali, cui appartengono anche le

---

<sup>141</sup> M. Couzinet, *De quelques observations quant à l'avis du Conseil constitutionnel sur la loi du 16 juin 1999 relative à la sécurité routière, et sur l'avis donné par le Conseil d'Etat le 27-09-1999 au Tribunal administratif de Versailles*, *Gaz. Pal.*, 26-27 juill. 2000, p. 4.

<sup>142</sup> Y. Mayaud, *op. cit.*, il quale esclude la possibilità di qualificare tale sanzione come amministrativo-punitiva, atteso che si tratta di un reato il cui accertamento è rimesso alla giurisdizione repressiva e non amministrativa; L. Desessard, *Le traitement simplifié des infractions*, *op. cit.*; J. Buisson, *obs. Crim 17 févr. 2004*, *Procédures*, n. 5, mai 2004, *Commentaire n. 112*; F. Ludwiczak, *Responsabilité pécuniaire de l'art. L. 121-3 du code de la route : de l'incertitude à l'imprécision*, *Rev. Pénitentiaire*, 2013, p. 547-554, in particolare p. 550, il quale però sottolinea il carattere ambiguo di tale responsabilità pecuniaria dotata di una finalità sicuramente punitiva e non riparatoria che la accosta alla responsabilità penale, mentre la giurisprudenza ne ha negato più volte tale natura (v. *infra* in nota). Per una sintesi delle opinioni, si veda J. P. Céré, *L'illusion pénal de la responsabilité du titulaire du certificat d'immatriculation*, *AJ Pén.*, sept. 2019, pp. 418-421.

<sup>143</sup> F. Ludwiczak, *op. cit.*

<sup>144</sup> Y. Mayaud, *op. cit.*; G. Defrance, *Le propriétaire du véhicule redevable pénalement*, *Jur. Aut.*, 2003, pp. 484-487; Ludwiczak, *op. cit.*, il quale nota l'atteggiamento ondivago della giurisprudenza penale, che tende a restare ancorata al dato letterale, così escludendo la natura penale della responsabilità (*supra* in nota).

<sup>145</sup> « il n'est certes par contestable que la disposition critiquée présente, indépendamment de ses finalités préventives et dissuasives, un aspect punitif. Il n'est donc pas douteux que sa conformité à la Constitution doit être vérifiée au regard des règles qui en découlent en matière répressive (...) De même est-il certain que, bien que n'instaurant pas une responsabilité pénale, au sens plein de ce terme, la mesure en cause est soumise, pour les mêmes raisons, au principe de présomption d'innocence », *Décision n° 99-411 DC du 16 juin 1999 - Observations du gouvernement*, [www.conseilconstitutionnel.fr](http://www.conseilconstitutionnel.fr).

contravvenzioni, per le quali però, come il Conseil ha cura di precisare, delle eccezioni al suddetto principio possono ammettersi.

La giurisprudenza è invece di contrario avviso e, sposando una tesi ‘formalistica’, si limita ad affermare quanto espressamente previsto dalla norma, cioè che il titolare del certificato di immatricolazione non è penalmente responsabile, senza però interrogarsi ulteriormente sulla natura di tale responsabilità<sup>146</sup>.

Resta allora aperta la questione della natura giuridica di questo tipo di responsabilità, per la quale una dottrina, negandone il carattere penale, sostiene che l’alternativa sarebbe il riconoscimento della sua natura civile, in quanto si tratterebbe di una forma di responsabilità assimilabile a quella del custode. Tuttavia, una simile conclusione è ritenuta difficilmente condivisibile alla luce degli argomenti adottati a sostegno della natura penale di tale responsabilità che, tra l’altro, non mira a riparare un danno, ma a sanzionare un’infrazione delle norme del codice della strada<sup>147</sup>.

Sposando la tesi che sostiene la natura penale della sanzione, si apre però un contrasto col principio della presunzione di innocenza vigente in materia penale; ciononostante essa è tollerata dall’ordinamento perché è posta in materia contravvenzionale, è verosimile (la titolarità del certificato di immatricolazione rinvia a un’alta probabilità di causazione del fatto da parte del suo titolare) ed è ammessa la prova contraria<sup>148</sup>. Tale soluzione era già stata condivisa anche dalla Corte Edu nell’*affaire Salabiaku c. France* del 7 ottobre 1988<sup>149</sup>, ove si afferma che gli ordinamenti dei singoli Stati conoscono e possono conoscere delle presunzioni di fatto o di diritto e, sebbene la Convenzione non le vieti, in

---

<sup>146</sup> Cass. crim., 31 janv. 2007, n. 07-81.365, Dr. Pén. 2007, comm. 138, obs. J.H. Robert; Cass. crim., 15 sept. 2010, pourvoi n. 09-877.326, che ha cassato la decisione della Corte d’appello che non aveva applicato l’art. 121-3 c. route così come modificato dalla legge del 12 maggio 2009, in quanto a seguito di tale modifica essa rappresentava una norma più severa. La Cassazione francese ha invece affermato che la sanzione pecuniaria non avesse natura di pena e per questo poteva essere applicata retroattivamente; Cass. crim., 5 déc. 2007, *Jur. Auto.*, 2008, p. 92; Cass. crim. 1 oct. 2003, *Jur. Auto.*, 2003, p. 626.

<sup>147</sup> Ludwiczak, *op. cit.*, pp. 552-553.

<sup>148</sup> G. Defrance, *L’application de l’article L. 121-3 du code de la route*, *Jur. Auto.*, nov. 2008, p.635.

<sup>149</sup> <https://hudoc.echr.coe.int>. Brevemente, questi i fatti: il signor Salabiaku veniva processato e condannato per il delitto di contrabbando di merci proibite (nella specie cannabis) punito dall’art. 392 par. 1 cod. des douanes, che prevede una presunzione di responsabilità. Il Salabiaku si rivolgeva ai giudici di Strasburgo lamentando, tra l’altro, la violazione dell’art. 6 par. 2 CEDU che sancisce la presunzione di innocenza. La Corte rigetterà il ricorso sottolineando che la presunzione non conosce un carattere assoluto nell’applicazione fattane dalla giurisprudenza, la quale ultima ammette la prova della forza maggiore a carico del pervenuto (par. 29). A parere della Corte, nel caso concreto la condanna riportata dal Salabiaku era compatibile con la presunzione di innocenza poiché il giudice nazionale non aveva applicato in modo automatico la norma censurata, ma nei due gradi di giudizio si era proceduto a un accertamento in concreto relativo non solo alla componente materiale del delitto, ma anche all’elemento intenzionale, nonostante tale ultimo accertamento non fosse richiesto dalla legge francese (par. 30).

materia penale obbliga gli Stati contraenti a prevederle entro limiti ragionevoli e in modo da garantire il diritto di difesa affinché esse non siano contrarie alla presunzione di innocenza di cui all'art. 6 CEDU<sup>150</sup>. La Corte di Cassazione francese non ha così sollevato ulteriori dubbi sulla compatibilità di tali presunzioni con la Convenzione<sup>151</sup> e pure la dottrina sembra tollerarle finché limitate alla materia contravvenzionale<sup>152</sup>, cioè con riguardo a comportamenti privi di un effettivo disvalore, ma per i quali si pone l'esigenza di individuare un responsabile.

Da ultimo, la natura penale comporta che questa forma di responsabilità pecuniaria è di stretta interpretazione e non può essere estesa analogicamente<sup>153</sup>, ma al più solo per mano del legislatore. Ed è quello che è accaduto con la riforma del 2021, che ha posto a carico del produttore una responsabilità pecuniaria che presenta molti elementi in comune con quella della norma appena analizzata.

Va però osservato che nell'alternativa tra natura penale e civile di questa forma di responsabilità, si palesa anche una terza via, volta a riconoscere il carattere amministrativo di tali sanzioni. Questa qualificazione è stata esclusa in dottrina per la sanzione in esame<sup>154</sup>, ma va osservato che la questione non è circoscritta alla materia del codice della strada, e si inserisce, piuttosto, nel più ampio dibattito, accennato nelle prime pagine del presente lavoro, sulla natura para-penale delle contravvenzioni in generale. Ebbene, a considerare le sanzioni di cui si è ora discusso come sanzioni amministrative punitive, potrebbe poi porsi la questione dell'estensione delle garanzie della CEDU, ove ritenute sostanzialmente penali secondo i criteri Engel; va però osservato che la dottrina francese ritiene che il carattere amministrativo possa essere riconosciuto solo a seguito di una depenalizzazione, mentre fintantoché permane la struttura tripartita delle infrazioni, non può che concludersi per la natura penale delle contravvenzioni, seppur con le peculiarità che le caratterizzano<sup>155</sup>.

---

<sup>150</sup> Il principio di diritto viene ribadito anche nell'affaire *Pham Hoang c. France*, Corte EDU, 25 settembre 1992. Sulla questione e, più in generale sulle presunzioni di colpa o di responsabilità, si rinvia a F. Desportes, F. Le Guehec, *op. cit.*, pp. 431 ss.

<sup>151</sup> Cass. Crim., 6 nov., 1991 : Juris Data n. 1991-003617; Bull. crim. 1991, n. 397; Cass. Crim., 9 apr. 1992: Juris data n. 1992-001931; Bull. Crim. 1992, n. 155.

<sup>152</sup> « *Les contraventions nous semblent s'accommoder parfaitement de la solution. Tout en ressortissant au droit pénal, les contraventions n'engagent pas à proprement parler la criminalité ou la délinquance. Elles ne sont que des atteintes à l'organisation de la vie sociale, contrairement aux crimes et aux délits, qui, par leur gravité intrinsèque, touchent aux valeurs essentielles de la société* », Y. Mayaud, *cit.*

<sup>153</sup> G. Defrance, *L'application de l'article L. 121-3 du code de la route*, *Jur. Auto.*, nov. 2008, p.633-641.

<sup>154</sup> È questa l'opinione, già ricordata *supra*, in nota, di Y. Mayaud, *op. cit.*

<sup>155</sup> Sul punto, si rinvia ancora a E. Dreyer, *Droit Pénal, cit.*, pp. 80 ss.

### 5.3 La lettura della norma sulla responsabilità del produttore alla luce della pronuncia del *Conseil constitutionnel* del 1999

Si è già avuto modo di notare la sovrapposibilità del nuovo art. 123-2 c. route sulla responsabilità del produttore del sistema di guida autonoma con l'art. 121-3 c. route, sulla responsabilità del soggetto titolare del certificato di immatricolazione.

Sulla base delle opinioni avanzate dalla dottrina con riferimento a quest'ultima disposizione, si ripropongono qui le stesse problematiche, la prima delle quali attiene all'oggetto della presunzione. In ragione della presunzione che già avvolge l'*élément moral* delle contravvenzioni, deve ritenersi che quest'altra presunzione interessi piuttosto l'elemento materiale, sicché in sostanza si riterrebbe provato l'apporto causale del produttore nella realizzazione dell'infrazione allorché la gestione delle operazioni di guida sia affidata al sistema di pilota automatico. Si tratterebbe, quindi di una *présomption de responsabilité* che si innesta su una già condivisa *présomption de culpabilité* in materia contravvenzionale.

L'introduzione di una nuova presunzione di responsabilità, a carico di un soggetto diverso dal proprietario, pone poi naturalmente il problema della compatibilità col principio di responsabilità per fatto proprio, sempreché a questa forma di responsabilità sia riconosciuta natura penale.

Orbene, potrebbe ritenersi che poiché si fa riferimento alla nozione di "*redevabilité*", già conosciuta nel sistema del code de la route all'art. 121-3, e poiché la dottrina concorda nel ritenere la natura penale di tale disposizione, allora debba concludersi che anche la norma sulla responsabilità del produttore ha natura penale. E, infatti, anche qui ricorrono quegli argomenti valorizzati dalla dottrina per sostenere la natura penale della sanzione posta a carico del titolare del certificato di immatricolazione e, in particolare, anche qui la sanzione irrogata è l'ammenda, che presuppone la commissione di un reato, nella specie una contravvenzione. Questi argomenti, che inducono la dottrina a non dubitare della natura penale della sanzione a carico del proprietario del veicolo, ci portano a leggere la sanzione pecuniaria posta a carico del produttore del sistema di intelligenza artificiale come una sanzione di natura penale, punitiva o quasi penale. Inoltre, il fatto che la norma non escluda espressamente il carattere penale di tale responsabilità sembra quasi una presa d'atto da parte del legislatore delle conclusioni raggiunte dalla dottrina intorno alla

disposizione dell'art. 121-3 c. route e delle contraddizioni che quella norma ha prodotto, con riferimento soprattutto ad alcune decisioni della giurisprudenza.

Ma a condividere questa conclusione, si pone il problema della compatibilità di tale forma di responsabilità penale col principio della presunzione di innocenza.

Come detto poco sopra, la questione si era già presentata per le norme del vecchio code de la route che introducevano la responsabilità del titolare del certificato di immatricolazione, ma era stata superata dal Conseil constitutionnel proprio facendo leva su tre argomentazioni, in ordine: l'ammissibilità della prova liberatoria gravante sul titolare di tale certificato; l'esistenza comunque di una *faute personnelle* laddove la prova liberatoria non fosse raggiunta; il dato normativo che esclude il carattere penale della responsabilità.

Nel caso in esame, non vi è alcun riferimento alla prova liberatoria e anzi il richiamo all'art. 121-3 c. route sarebbe in parte irragionevole poiché si tratta di prove incompatibili con la posizione del produttore; conseguentemente, non vi sarebbe *faute personnelle* poiché essa postula l'ammissibilità della prova contraria. Infine, la norma non esclude testualmente il carattere penale di tale forma di responsabilità. Sembra, allora, che questo obbligo pecuniario possa entrare in conflitto col principio di personalità della responsabilità penale e con la presunzione di innocenza, posto che non sono presenti quei requisiti enunciati nel 1999 dalla sentenza del Conseil constitutionnel in presenza dei quali tale conflitto con due principi fondamentali dell'ordinamento francese non si verifica.

La dottrina che ha commentato l'introduzione della nuova disposizione non ha espresso alcuna perplessità sul punto e anzi ha richiamato proprio la *décision* n. 99-411 del 16 giugno 1999 per ricordare che tali forme di responsabilità sono compatibili con la Costituzione<sup>156</sup>.

Volendo argomentare alla luce di questa impostazione che sottrae la nuova disposizione da censure di incostituzionalità, può osservarsi che comunque il produttore ha la possibilità di addurre delle prove a sua difesa e, nella specie, può provare che il sistema non aveva il controllo dinamico o che lo stesso non era utilizzato in conformità con le istruzioni fornite.

---

<sup>156</sup> S. Detraz, *op. cit.*,

Al più potrebbe sostenersi il richiamo, in via analogica, della prova contraria di cui all'art. 121-3 c. route laddove fa riferimento all'evento di forza maggiore (e non anche al furto -prova che, come detto, è logicamente incompatibile con la posizione del produttore- o all'individuazione di un soggetto terzo autore dell'infrazione).

In ultima analisi, va evidenziato che la presunzione assoluta a carico del produttore può essere evitata ammettendo la prova della sopravvenienza di una causa di forza maggiore alla luce dei principi generali dell'ordinamento penale francese, in particolare dell'art. 121-3, comma 5 c. pén. che, proprio con riferimento alle contravvenzioni, afferma che: “*il n'y a point de contravention en cas de force majeure*”. Ma in proposito non va sottaciuto che la giurisprudenza della Cour de cassation è molto restia a riconoscere cause di forza maggiore<sup>157</sup>, soprattutto quando si tratti di guasti del veicolo, anche allorché tali guasti siano causati da fattori indipendenti dalla condotta del conducente<sup>158</sup>. E non è tutto.

Il Conseil constitutionnel sottolinea che, laddove la prova liberatoria non fosse raggiunta, vi deve comunque essere una *faute personnelle*; in questo caso la colpa del produttore consisterebbe nella mancata progettazione di un sistema esente da vizi o da difetti che ne hanno comportato il malfunzionamento. Nondimeno, anche a voler dimostrare che la violazione della regola posta dal codice della strada si è verificata per via di un difetto che in fase di produzione si sarebbe potuto evitare, è vero che la responsabilità potrebbe non essere del produttore, ma di soggetti che hanno contribuito alla realizzazione di parti del sistema che si siano rivelate fallaci o difettose. L'isolamento del singolo comportamento colposo nella catena di produzione del sistema non è cosa semplice: più sono gli attori e più l'errore è difficile da rintracciare<sup>159</sup>. È vero che proprio questa difficoltà giustifica l'instaurazione di un sistema di presunzione di responsabilità, in cui la prova liberatoria è posta a carico di colui che dispone di tutte le informazioni sulla progettazione e costruzione del veicolo, molte delle quali potrebbero essere protette dal segreto industriale<sup>160</sup>; ma è anche vero che le presunzioni devono essere ragionevoli, oltre che non assolute e che, mentre appaiono giustificate in materia di responsabilità

---

<sup>157</sup> Per una sintesi della casistica della giurisprudenza, si rinvia a R. Josseaume, *La force majeure en droit routier: un mirage législatif?*, *Gaz. Pal.*, avr. 2018, n. 16, pp. 19-21.

<sup>158</sup> Non è stata ritenuta causa di forza maggiore la presenza di acqua che causava il guasto dell'indicatore di cambio di direzione di un furgone guidato da un dipendente che prendeva servizio al mattino e che non aveva condotto lo stesso veicolo i giorni prima, Cass. crim. 6 nov. 2013, *Bull. Crim.* 215.

<sup>159</sup> M. Monot-Fouletier, *Véhicule autonome: vers une autonomie du régime de responsabilité applicable ?*, *Recueil Dalloz*, 2018, p. 129.

<sup>160</sup> M. Monot-Fouletier, *op. cit.*

civile, attesa la finalità riparatoria della sfera giuridica della vittima<sup>161</sup>, lo sono meno in un contesto, quale quello penale, in cui la finalità è punitiva. È quindi ragionevole addossare al produttore l'obbligo di pagare l'ammenda senza possibilità di rivalsa (ad esempio nei confronti di produttori terzi che gli hanno fornito componenti difettose) e fatta salva la sola prova della forza maggiore?

Sembra allora che una vera *faute* al produttore non possa essere riconosciuta e la presunzione di nuova introduzione potrebbe porre dei dubbi di compatibilità col principio di presunzione di innocenza. La domanda è al momento aperta e costituirà probabilmente oggetto di dibattito in dottrina, nell'attesa delle prime applicazioni che di essa farà la giurisprudenza.

Bisogna comunque sempre tenere a mente che questi quesiti hanno una portata attenuata perché si pongono in un contesto contravvenzionale, quindi di infrazioni penali "minori", che per molti aspetti si avvicinano alle sanzioni amministrative e per le quali l'elemento morale è genericamente presunto, come è stato brevemente osservato nella prima parte di questo lavoro. Per questo la presunzione di responsabilità sollecita meno preoccupazioni, almeno finché limitata alla sfera delle contravvenzioni.

## **6. Un esempio da imitare?**

Si conclude qui l'analisi della riforma del legislatore francese, la prima ad essere approvata da uno Stato membro dell'Unione europea. A questo punto possiamo chiederci se a fronte delle nuove sfide poste dall'intelligenza artificiale al legislatore del XXI secolo, quello francese rappresenti un tentativo ben riuscito di disciplina di un settore emergente quale è quello delle auto a guida autonoma.

Come già evidenziato si tratta di un testo normativo che ha una portata allo stato più teorica che pratica: fintantoché le auto a guida delegata non inizieranno a circolare, queste norme non potranno trovare applicazione.

L'analisi qui offerta ha messo in luce i vari punti problematici, rappresentati anzitutto dall'individuazione della reale portata innovatrice della norma sulla responsabilità del conducente che sembra piuttosto dettare i requisiti che un sistema di guida autonoma deve possedere per poter essere messo in commercio e che probabilmente si concentra troppo poco sulle regole cautelari destinate a orientare il comportamento del conducente.

---

<sup>161</sup> *Ibidem.*

Ancora, un punto debole della riforma è l'impiego di una nozione generica di "produttore", che tralascia la posizione di soggetti aventi un ruolo fondamentale nella catena produttiva, quali l'ideatore e l'addestratore dell'algoritmo. Le conseguenze non sono irrilevanti soprattutto in un settore, quello penale, in cui è fatto divieto di ricorrere al meccanismo analogico *in malam partem*. Il testo, pertanto, appare poco attento nell'individuare il contesto produttivo di tali sistemi e a isolare le figure sulle quali può ricadere la responsabilità. Tale lacunosità è probabilmente il risultato dell'assenza di un dibattito parlamentare prima dell'entrata in vigore tale riforma che, come più volte ricordato, è stata adottata con la veste formale di una *ordonnance* e che, se sarà ratificata dal Parlamento, probabilmente non subirà alcuna modifica.

In ogni caso, è anche vero che le perplessità suscitate da questi vuoti di disciplina si dissolvono facendo ricorso all'applicazione delle regole generali del code pénal che, come si è visto, prevalgono su quelle del code de la route o si affiancano alle stesse in caso di lacune.

Si potrebbe a questo punto osservare che se il nuovo costringe a ricorrere al vecchio, cioè se la riforma necessita comunque di un'integrazione costante da parte delle regole generali del diritto penale, allora sembra svanire l'utilità della stessa, che perde i connotati di disciplina, per mostrarsi solo come atto politico, privo di una reale portata giuridica innovativa. Se ciò non può essere smentito, va però notato che la novella ha il merito di adattare le regole generali in tema di responsabilità penale a un nuovo contesto, quello delle auto a guida autonoma, appunto, che ha richiesto al legislatore di riflettere sulla posizione del conducente e su quella del produttore e di individuare il confine tra la responsabilità dell'uno e dell'altro, tracciata ricorrendo alla nozione di "controllo dinamico".

D'altra parte, però, questo conferma che i problemi derivanti dall'utilizzo dell'intelligenza artificiale nel quotidiano sono risolvibili con gli schemi ordinari propri del diritto penale e che, almeno per ora, discutere di una responsabilità del sistema di IA rappresenta solo un esercizio giuridico, perché non esistono sistemi che operano senza la supervisione dell'uomo.

Un altro aspetto che merita attenzione è il rischio reso evidente da questa riforma: trovare in ogni caso un responsabile. Questa è l'impressione che emerge dalla lettura della disposizione sulla responsabilità del produttore per le contravvenzioni previste dal code de la route. Se è vero che il conducente non può essere ritenuto responsabile per un "errore" del sistema, è anche vero che non sempre questo "errore" può essere addebitato al

produttore: responsabile sì, ma non in ogni caso. Come già osservato, il problema è attenuato dal fatto che si pone per i reati contravvenzionali in tema di circolazione stradale, cioè infrazioni che nel nostro ordinamento sono amministrative e per le quali, pertanto, non potrebbe essere sollevato un dibattito simile a quello che ha interessato in passato le disposizioni del code de la route che presentano una struttura del tutto sovrapponibile alla norma sulla responsabilità del produttore. Inoltre la realtà dei fatti potrà presentare situazioni nelle quali il produttore è del tutto disinteressato a fornire la prova contraria, perché magari riterrà più conveniente versare l'importo dell'ammenda piuttosto che sostenere le spese di un giudizio.

Probabilmente questa scelta è anche ispirata da ragioni economiche e di promozione dei prodotti intelligenti: porre a carico del produttore il pagamento dell'ammenda può incentivare le vendite e scongiurare la ritrosia dei consumatori verso le nuove tecnologie. È un ragionamento che segue una logica di mercato, ma è bene sottolineare che queste dinamiche non possono influenzare le scelte giuridiche del legislatore penale, al quale non è consentito sacrificare il principio della responsabilità personale per seguire ragioni economiche. Si tratta di un bilanciamento che nel contesto della riforma francese può essere tollerato in ragione della natura para-penale delle contravvenzioni, ma che non va esteso a sanzioni strettamente penali, mentre potrebbe essere ammesso in sede civile, dove si può immaginare la valorizzazione da parte del legislatore di forme di risarcimento o di indennizzo come alternativa alla scelta di creare aree di immunità per il produttore.

Il quadro giuridico in materia di responsabilità civile potrà allora costituire la sede in cui si bilanceranno le esigenze sull'affidabilità, sicurezza e fiducia dei nuovi prodotti e servizi, l'efficace tutela delle potenziali vittime di danni o pregiudizi e lo sviluppo delle nuove tecnologie e prodotti, come riconosce lo stesso legislatore europeo nella Risoluzione dell'ottobre 2020<sup>162</sup>; per contro, il diritto penale non potrà prestarsi a logiche di bilanciamento di questo tipo, e ogni accertamento di responsabilità dovrà mostrarsi rispettoso dei principi fondamentali che stanno alla base degli ordinamenti moderni.

Concludendo, possiamo a questo punto interrogarci sull'*opportunità* di tale riforma, riflettendo anzitutto sulla *necessità* di tale intervento da parte del legislatore. La risposta è nel senso di un intervento non necessario se si considera che auto di questo tipo non

---

<sup>162</sup> Si veda il considerando di cui alla lettera B della Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale (2020/2014(INL)).

sono ancora in circolazione, quindi la riforma serve semmai ad incoraggiarne la produzione.

In secondo luogo, potremmo domandarci se questa riforma era *indispensabile*. Anche a tale domanda potremmo dare una risposta negativa, posto che si tratta di una riforma quasi “di facciata”, che richiede sotto più aspetti il supporto degli schemi ordinari del diritto penale e che sul riparto di responsabilità giunge a conclusioni cui già le regole generali del code pénal avrebbero condotto.

A questo punto il quesito principale ruota attorno all'*utilità* di tale intervento normativo e sarebbe avventato ricavare la risposta in modo affrettato dalle due precedenti. Una più attenta riflessione consente di riconoscere come la riforma rappresenti uno strumento *utile* per l'interprete in un contesto di responsabilità delicato, qual è quello penale. E infatti, le nuove norme:

a) forniscono la definizione di “controllo dinamico”, segnando così un confine chiaro tra la responsabilità del produttore e quella del conducente, confine che altrimenti sarebbe spettato all'interprete individuare;

b) istituiscono per via legale una posizione di garanzia in capo al conducente che è chiamato a riprendere in mano il controllo della vettura su richiesta del sistema, posizione di garanzia sulla quale altrimenti si sarebbe potuto dubitare;

c) esplicitando dei doveri di condotta, incidono sulla qualificazione della colpa, che, per usare termini noti alla dottrina italiana, da “generica” (*faute simple*), diviene “specifica” (*faute délibérée*), con le conseguenze che da ciò discendono sul piano probatorio.

In conclusione, con tale intervento normativo il legislatore francese fornisce il proprio contributo alla concreta attuazione del principio di prevedibilità delle conseguenze sanzionatorie della norma penale, che a livello convenzionale trova base nell'art. 7 della CEDU e che è un obiettivo da perseguire da parte dello Stato e dei suoi organi legislativi e giurisprudenziali a tutela dei diritti dell'individuo e delle garanzie che il diritto penale è chiamato a riconoscergli<sup>163</sup>.

---

<sup>163</sup> G. Grasso, F. Giuffrida, *Gli effetti della giurisprudenza della Corte Edu sull'ordinamento italiano: prospettive di diritto penale sostanziale*, p. 61, in *L'indice penale*, Gennaio-aprile 2020.



## RIFLESSIONI CONCLUSIVE

Quell'interlocutore al quale spiego l'argomento oggetto della mia tesi, che ho presentato al lettore nell'Introduzione, attende sempre da parte mia una risposta alla domanda "chi è responsabile se una condotta penalmente rilevante è tenuta durante l'uso di un sistema di IA?". Purtroppo, io, una risposta certa non ho mai saputo darla e probabilmente nessuno allo stato attuale può offrirmi una.

Ci ha provato il legislatore francese, con la riforma analizzata nel CAPITOLO IV, con riguardo al ristretto ambito dei reati commessi durante l'utilizzo di auto dotate del sistema di guida assistita. È vero: la riforma non va esente da critiche, ciononostante rappresenta un primo punto fermo in una materia nuova che ha bisogno di riferimenti di natura legislativa.

Dalla presente indagine emerge che le categorie tradizionali del diritto penale sono allo stato idonee e sufficienti a fronteggiare i nuovi rischi e che la creazione di una responsabilità *singolare* dei sistemi di IA sul piano penale, oltre che apparire difficilmente praticabile per le ragioni esposte nell'introduzione al CAPITOLO III, non si rende necessaria.

In particolare, si è visto nella SEZIONE I dello stesso capitolo come, sul piano dell'illiceità colposa, la presenza di regole cautelari può essere utile per orientare la responsabilità tra la persona del produttore e quella dell'utente. Ciò che è richiesto al legislatore, pertanto, è anzitutto un sistema di regole che indirizzi la produzione dei sistemi di IA e che miri a prevenire i danni e non solo a porvi rimedio. Le regole sulla produzione serviranno allora a incrementare la sicurezza dei prodotti intelligenti. Certo, dovrà trattarsi di regole sulla produzione e non di pastoie per la produzione e questo sarà un compito molto delicato, di cui già il legislatore europeo si sta occupando, probabilmente talvolta scegliendo di perseguire la via di una legislazione "simbolica", inattuabile, come quando si illude di neutralizzare i pericoli connessi ai sistemi ad alto rischio richiedendo che il loro addestramento sia basato su dati "esenti da errori". Di questo si è già discusso nel CAPITOLO II, § 1.5, dando atto del fatto che dati esenti da errori e sistemi a rischio zero non esistono. Il requisito di dati privi di errori non può ragionevolmente essere imposto se consideriamo che il volume di dati prodotti nel mondo raggiunge livelli elevatissimi, tanto è vero che per quantificarli è stata coniata la nuova misura di *zettabyte*, pari a un trilione di *gigabyte* (ovvero 250.000.000.000 DVD!) e che si stima che, se nell'anno 2018 il volume totale di dati creati nel mondo è stato di 28

zettabyte (ZB), registrando un aumento di più di dieci volte rispetto al 2011, entro il 2025 il volume complessivo di dati arriverà fino a 163 ZB<sup>1</sup>.

Accanto a una normazione preventiva, che sia però ragionevole, è emersa l'esigenza di un ripensamento dello strumento penale, al fine di introdurre nuove fattispecie incriminatrici che prediligano la struttura di reati di pericolo piuttosto che quella di reati di danno, per i quali la prova del nesso causale talvolta sarà molto difficile da raggiungere, come si è tentato di spiegare nella SEZIONE II del CAPITOLO III. Su questo punto va però tenuto presente che quando il legislatore maneggerà lo strumento penale non dovrà farlo con l'idea di trovare in ogni caso un responsabile. Questo ad esempio è quanto sembra aver fatto il legislatore francese, chiamando a rispondere per le contravvenzioni consistenti nella violazione delle norme del codice della strada il produttore dell'auto a guida delegata. Se questo è vero, è però anche da sottolineare la natura para-penale di tali contravvenzioni, che nel nostro ordinamento hanno natura amministrativa. In ragione di ciò, l'intervento del legislatore francese può ritenersi condivisibile in una materia che ha più carattere amministrativo, che penale, mentre simili forme di responsabilità non possono trovare spazio nella materia sostanzialmente penale.

Queste riflessioni ci portano ad affermare che la sfida posta dalle tecnologie di IA mette alla prova non solo le categorie tradizionali del diritto penale, anche d'evento, ma, prima ancora, i principi fondamentali dell'ordinamento penale. Tra questi, v'è il principio della responsabilità penale *personale*, che richiede la prova, oltre ogni ragionevole dubbio, degli elementi del reato, quindi anche della causalità e della colpevolezza. La responsabilità non può che essere per un *fatto proprio e colpevole*.

La colpevolezza colposa ha ispirato gli approfondimenti proposti nella SEZIONE III del CAPITOLO III, dedicato alle peculiarità del caso concreto che possono assumere rilievo penale. Il tema è quello delle circostanze anomale che possono rendere inesigibile una condotta conforme a diligenza da parte dell'agente concreto, escludendo la colpevolezza, ovvero possono agire in un momento anteriore, escludendo addirittura il reato in quanto tale, con conseguente assoluzione "perché il fatto non costituisce reato", ove siano tali da integrare un'ipotesi di caso fortuito o forza maggiore. Si tratta di situazioni anomale che, verosimilmente, si paleseranno durante l'uso del sistema da parte dell'utente e che,

---

<sup>1</sup> Dati tratti dal rapporto finale dell'indagine conoscitiva sui Big Data condotta congiuntamente dall'Autorità per le Garanzie nelle Comunicazioni, dall'Autorità Garante della Concorrenza e del Mercato e dal Garante per la Protezione dei Dati Personali, pubblicata il 10.02.2020, <https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9264297>, pp. 5-6.

pertanto, interesseranno maggiormente la posizione di questi piuttosto che quella del produttore.

In questo tracciato di responsabilità, l'utente riveste un ruolo tutt'altro che marginale. Il suo dovere di diligenza si compone del dovere di informazione, vigilanza e intervento; valutando la sua condotta alla luce del parametro dell'agente modello, spesso si sarà portati a concludere che l'utente ideale avrebbe saputo gestire quel rischio verificatosi nel caso di specie. Nondimeno, ciò non può ritenersi sufficiente per affermarne la responsabilità penale. Le situazioni in cui si sarà chiamati a giudicare il comportamento dell'utente costituiranno un importante momento di riflessione e di valorizzazione della colpevolezza colposa, fase dell'illecito in cui bisogna prendere in considerazione gli elementi utili per un'ulteriore personalizzazione del giudizio. Così, se si riterrà che l'agente modello avrebbe avuto una prontezza di riflessi che gli avrebbe permesso di riprendere il controllo della vettura che, non riconoscendo un segnale di stop, proseguiva la marcia investendo un pedone, potrà essere provato che una simile reazione salvifica era inesigibile da parte del conducente in quella peculiare situazione. Dovranno allora essere finanziati e sostenuti quegli studi, di cui si è dato atto nella SEZIONE III, § 7, CAPITOLO III, che già dimostrano come la capacità di riprendere il controllo del volante sia fortemente ridotta quando il conducente non ha un ruolo attivo alla guida; di essi il diritto penale dovrà avvalersi ai fini del giudizio di esigibilità della condotta doverosa. Allo stesso modo, andrà favorita la ricerca di sistemi di IA che siano interpretabili o spiegabili, al fine di consentire l'accertamento dell'eventuale errore algoritmico in sede processuale.

Ben si comprende, allora, come l'intelligenza artificiale è un prodotto della scienza ma, al contempo, una sfida ancora attuale per la scienza: la scienza informatica, chiamata a creare sistemi più sicuri e interpretabili; la scienza che studia le interazioni uomo-macchina, per comprendere quali siano le responsabilità dell'utente e ciò che è da questi esigibile; il diritto che, in quanto scienza, è chiamato a dare le sue risposte sul piano della prevenzione e della sanzione. Il diritto andrà però considerato nel suo complesso. Probabilmente in un contesto di incertezza così elevata, dovrà accettarsi, da parte del giurista, ma anche della collettività, un ruolo marginale del diritto penale, affinché il progresso tecnologico sia conferma delle conquiste della civiltà giuridica moderna e non un contesto favorevole all'emersione di istanze securitarie che affidano un ruolo cardine alla sanzione penale. ad emergere in modo evidente da questo studio sono i limiti del diritto penale di fronte ai problemi della quarta rivoluzione, la quale fa sorgere istanze di

tutela che non si può pensare di soddisfare interamente attraverso il ricorso allo strumento più forte che il potere statale possiede e che si spinge sino al punto di privare l'uomo di uno dei suoi valori più grandi, la libertà. Come già ammoniva Federico Stella nei primi anni duemila, “la soluzione dei problemi della modernità non può essere affidata in via esclusiva al diritto penale”; occorre piuttosto un incontro sinergico tra tutte le discipline giuridiche, che valorizzi i diversi aspetti nei quali si articola la tutela della vittima e costruisca risposte più adeguate alle sfide della modernità<sup>2</sup>.

In quest'ottica, il diritto penale è e dovrà rimanere l'ultimo rimedio, l'*extrema ratio*. Ecco, questa penso possa essere la risposta che certamente potrò dare all'interlocutore che mi domanderà quale sia la mia opinione sul mio tema di ricerca sull'Intelligenza Artificiale e la responsabilità penale. Un punto di partenza, piuttosto che una conclusione.

---

<sup>2</sup> F. Stella, *Giustizia e modernità*, cit., p. 16.



## BIBLIOGRAFIA

AA.VV., *Macchine che pensano. La nuova era dell'intelligenza artificiale*, New scientist, Edizioni Dedalo, Bari, 2018.

Abbot, R., *The Reasonable Robot. Artificial Intelligence and the Law*, Cambridge, Cambridge University Press, 2020

Agazzi, E., *La spiegazione causale di eventi individuali (o singoli)*, in *Riv. it. dir. proc. pen.*, 1999, pp. 393 ss.

Alpaydin, E., *Introduction to Machine Learning*, The MIT Press, Cambridge, 2010.

Amato, S., *Biodiritto 4.0. Intelligenza artificiale e nuove tecnologie*, Torino, Giappichelli, 2020,

A. Amidei, *Robotica intelligente e responsabilità: profili e prospettive evolutive del quadro normativo europeo*, in *Intelligenza artificiale*, cit. p. 86.

Amidei, A., *Intelligenza artificiale e diritto – Intelligenza artificiale e product liability: sviluppi del diritto dell'Unione europea*, in *Giur. it.*, 2019, 7, p. 1715 ss.

Anderson, M., S.L. Anderson (a cura di), *Machine Ethics*, Cambridge, Cambridge University Press, 2011.

Angioni, F., *Note sull'imputazione dell'evento colposo con particolare riferimento all'attività medica*, in *Studi in onore di Giorgio Marinucci*, Milano, Giuffrè, 2006, p. 1279 ss

Angwin, J., Larson, J., Mattu, S., Kirchner, L., *Machine Bias*, in [www.propublica.org](http://www.propublica.org), 23 maggio 2016.

Antolisei, F., *Il rapporto di causalità nel diritto penale*, Padova, CEDAM, 1934.

Athanasidou, E., Geradts, Z., Van Eijk, E., *Camera Recognition with Deep Learning*, 3 *Forensic Sci. Res.* 210 (2018).

Balbi, G., *La volontà e il rischio penale d'azione*, Napoli, Jovene 1995.

Balkin, J. M., *The Three Laws of Robotics in the Age of Big Data*, 78 *Ohio St. L.J.* (Sep. 10, 2017)

Barocas, S., Selbst, A. D., *Big Data's Disparate Impact*, in *California Law Review*, Vol. 104, pp. 671 ss.

Bartoli, R., *Colpevolezza: tra personalismo e prevenzione*, Torino, Giappichelli, 2005.

- Bartoli, R., *Il problema della causalità penale*, Torino, Giappichelli, 2010.
- Bartoli, R., *Diritto penale e prova scientifica*, in G. Canzio, L. Luparia (a cura di), *Prova scientifica e processo penale*, Padova, CEDAM, 2018, p. 86.
- Bartoli, R., *La responsabilità penale da esposizione dei lavoratori ad amianto. Un'indagine giurisprudenziale*, [www.penalecontemporaneo.it](http://www.penalecontemporaneo.it).
- Bartoli, R., voce *Fonti della colpa*, *Enc. dir., I Tematici, Il reato colposo*, Milano, Giuffrè, pp. 519 ss.
- Bartolucci, M.A., *L'art. 8 d.lgs. n. 231 del 2001 nel triangolo di Penrose, tra minimizzazione del rischio-reato d'impresa e "nuove forme" di colpevolezza*, *ibidem*, pp. 112-133.
- Basile, F., *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. e uomo*, 29 settembre 2019
- Bassini, M., Liguori, L., Pollicino, O., *Sistemi di intelligenza artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, in *Intelligenza artificiale, protezione dei dati personali e regolazione*, a cura di F. Pizzetti, 2018, p. 338.
- Beck, S., *Google cars, software agents, autonomous weapons systems – New challenges for criminal law?*, in E. Hilgendorf, U. Seidel (eds.), *Robotics, Autonomics, and the Law*, Baden-Baden, 2017, 227 ss.
- Belkin, M., Hsu, D., Ma, S., Mandal, S., *Reconciling modern machine learning practice and the classical bias-variance trade off*, in *PNAS*, vol. 116, n. 32, 2019, p. 15849-14854.
- Bénéjat, M., *Le droit pénal de la route face aux nouveaux modes de transport*, *AJ Pén.*, sept. 2019, pp. 428-430.
- Bifulco, R., *Intelligenza artificiale, internet e ordine spontaneo*, in F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018, p. 390.
- Bjerring, J. C., Busch, J., *Artificial Intelligence and Patient-Centered Decision-Making*, *Phil. and Tech*, 8 jan. 2020, pp. 349-371.
- Blaiotta, R., *Causalità e colpa: diritto civile e diritto penale si confrontano*, in *Cass. Pen.*, 1, 2009, pp. 78 ss.
- Blaiotta, R., *La ricostruzione del nesso causale nelle esposizioni professionali*, in *Cass. pen.*, 2, 2006, pp. 800 ss.
- Boden, M. A., *L'intelligenza artificiale*, Bologna, Il Mulino, 2018.
- Borruso, R., *Analisi giuridica del computer*, in *Computer e diritto*, tomo I, Giuffrè, Milano, 1988.

Borsari, R., *Intelligenza artificiale e responsabilità penale; prime considerazioni*, *Media Laws*, p. 265.

Bricola, F., *Responsabilità penale per il tipo e il modo di produzione*, in *La responsabilità dell'impresa per i danni all'ambiente e ai consumatori*, Milano, Giuffrè, 1978, pp. 75 ss.

Borsari, R., *Intelligenza Artificiale e responsabilità penale: prime considerazioni*, in *MediaLaws*.

Borsatti, E., *La soggettivazione della colpa: profili dogmatici e pratico-applicativi*, in *Ind. pen.* 2005, pp. 75-99

Burchard, C., *L'intelligenza artificiale come fine del diritto penale? sulla trasformazione algoritmica della società*, in *Riv. it. dir. proc. pen.*, 4, 2019, pp. 1909 ss.

Buisson, J., *Les Présomptions de culpabilité, Procédures – Ed. Juris-Classeur*, 12, 1999, pp. 3-4.

Burrell, J., *How the machine 'thinks': Understanding opacity in machine learning algorithms*. *Big Data & Society*, 3(1), 2016.

Cabrillac, R., *Introduction générale au droit*, Paris, Dalloz, 2021.

Calo, Ryan and Evtimov, Ivan and Fernandes, Earlene and Kohno, Tadayoshi and O'Hair, David, *Is Tricking a Robot Hacking?*, (March 27, 2018), *University of Washington School of Law Research Paper No. 2018-05*, Available at SSRN: <https://ssrn.com/abstract=3150530> or <http://dx.doi.org/10.2139/ssrn.3150530>

Canepa, A., *L'imputazione soggettiva della colpa. Il reato colposo come punto cruciale nel rapporto tra illecito e colpevolezza*, Torino, Giappichelli, 2011.

Canestrari, S., *La doppia misura della colpa nella struttura del reato colposo*, in *Ind. pen.*, 2012. Pp. 21-30

Canestrari, S., *La colpa*, in *Trattato di diritto penale, Parte generale, vol. II, Il Reato*, a cura di Cadoppi, A., Canestrari, S., Manna, A., Papa, M., Torino, UTET, 2013.

Capilli, G., *Responsabilità e robot*, in *Nuova giur. civ.*, 2019, 3, 621.

Capilli, G., *Responsabilità e robot*, in *Nuova Giur. Civ.*, 2019, 3, p. 621 ss.

Cappellini, A., *Machina delinquere potest? Brevi appunti sull'intelligenza artificiale e la responsabilità penale*, *Discrimen*, 27.03.2019.

Cardon, D., *Le Pouvoir des algorithmes*, in *Pouvoirs*, 164, 2018.

Carlucci Aiello, L., *Intelligenza artificiale*, in *Enciclopedia Italiana – IX Appendice*, 2015, [www.treccani.it](http://www.treccani.it).

Carnap, R., *The Logical Structure of the World: Pseudoproblems in Philosophy*, University of California Press, 1967.

Carobene, A., voce *Intelligenza artificiale – Approfondimento*, in *Enc. Treccani online*, [www.treccani.it](http://www.treccani.it).

Carter C., et al., *The Credit Card Market and Regulation: In Need of Repair*, 10 *N.C. Banking inst.* 23, 2006, p. 41.

Caselli, A., *Dagli artifici dell'intelligenza, all'Intelligenza Artificiale*, in *Intelligenza artificiale, protezione dei dati personali e regolazione*, F. Pizzatti (a cura di), Giappichelli, Torino, 2018, p. 192.

Castronuovo, D., *La colpa penale*, Milano, Giuffrè, 2009.

Castronuovo, D., *L'Evoluzione teorica della colpa penale tra dottrina e giurisprudenza*, in *Riv. it. dir. proc. pen.*, 2011, pp. 1594-1645.

Castronuovo, D., *La colpa "penale". Misura soggettiva e colpa grave*, in *Riv. it. dir. proc. pen.* 2013, pp. 1723 ss

Castronuovo, D., voce *Colpa penale*, *Enc. dir., I Tematici, Il reato colposo*, Milano, Giuffrè 2021, pp. 200 ss.

Céré, J. P., *L'illusion pénal de la responsabilité du titulaire du certificat d'immatriculation*, *AJ Pén.*, sept. 2019, pp. 418-421.

Chagal-Feferkorn, K., *The Reasonable Algorithm*, in *University of Illinois Journal of Law, Technology & Policy*, 2018, no. 1 (Spring 2018), pp. 111-148

Chan, L., *Explainable AI as Epistemic Representation*, in *AISB 2021 Symposium Proceedings: Overcoming Opacity in Machine Learning* (ed. C. Zednik, Boelsen).

Chander, A., *The Racist Algorithm?*, in *Michigan Law Review*, April 2017, p. 1034 ss.

Christodoulou, H., *Quand la responsabilité pénale embarque à bord d'un véhicule à délégation de conduite*, *Gaz. Pal.* n°24, 29 juin 2021, p. 10.

Citron D. K., Pasquale, F., *The Scored Society: Due Process for Automated Predictions*, in *University of Maryland Francis King Carey School of Law, Legal Studies Research Paper*.

Citron, D. K., *Technological Due Process*, 85 *Wash. U. L. Rev.* 1249, 1260–63 (2008).

Civello, G., *Prevedibilità e reato colposo*, in *Enc Dir., I Tematici, Il reato colposo*, Milano, Giuffrè, p. 1004 ss.

Colonna, K., *Autonomous Cars and Tort Liability*, 4 CASE W. RES. J. L. TECH & INTERNET 81, 102- 04 (2012).

Commaret, D., *La loi du 10 juillet 2000 et sa mise en œuvre par la chambre criminelle de la Cour de cassation*, Gaz. Pal, 12-13 avril 2002, p. 4.

Commaret, D., *La loi Fauchon, cinq ans après, Droit pénal n° 4, Avril 2006, étude 7.*

Commaret, D., *La responsabilité pénale des décideurs en matière de délits non intentionnels depuis la loi du 10 juillet 2000*, Gaz. Pal. N. 255, 11.09.2004, p. 3.

Consulich, F., *Il nastro di Mobius. Intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato*, in Banca, borsa, tit. credito, 2018, pp. 224 ss.

Consulich, F., *voce Rischio consentito, Enc. dir., I Tematici, Il reato colposo*, Milano, Giuffrè, 2021, pp. 1102 ss

Conte, Ph., *Le lapiste et la mort, Dr. Pén. n° 1, Janvier 2001, chron. 2.*

Conte, Ph., *obs. Cass. Crim, 9 avr. 2019, Dr. Pén. 2019, comm. 103.*

Conte, Ph., *obs. Cass. crim. 3 déc. 2019, n. 19-82492, Dr. Pén. 2020.*

Contissa, G, Lagioia, F., Sartor, G., *La manopola etica: i veicoli autonomi eticamente personalizzabili e il diritto, Sistemi intelligenti, III, 2017, pp. 601-14 .*

Copi, I. M., Cohen, C., *Introduzione alla logica*, Bologna, Il Mulino, 1997.

Cordeschi, R., D'Avanzo, E., *Nuove prospettive nell'intelligenza artificiale*, in Enc. Treccani, XXI Secolo, Norme e Idee, 2009, pp. 183 - 190.

Costanza, M., *L'intelligenza artificiale e gli stilemi della responsabilità civile*, in Giur. It., luglio 2019, p. 1687.

Cotta, S., *La sfida tecnologica*, Il Mulino, Bologna, 1968

Cotte, B., Guihal, D., *La loi Fauchon, cinq ans de mise en œuvre jurisprudentielle, Dr. pén. n° 4, Avril 2006, étude 6.*

Coulon, C., *Révision de la Convention de Vienne sur la circulation routière : les voiture autonomes (pas tout à fait) sur la ligne de départ, Resp. civ. et assurance, n. 6, alerte 7, 2016.*

Couzinet, M., *De quelques observations quant à l'avis du Conseil constitutionnel sur la loi du 16 juin 1999 relative à la sécurité routière, et sur l'avis donné par le Conseil d'Etat le 27-09-1999 au Tribunal administratif de Versailles*, Gaz. Pal., 26-27 juill. 2000, p. 4.

Craik, K., *The Nature of Explanation*, Cambridge University Press, 1943.

Creel, K. A., *Function and User-Satisfaction in Explainable AI*, in *AISB 2021 Symposium Proceedings: Overcoming Opacity in Machine Learning* (ed. C. Zednik, Boelsen).

Crespi, A., *Riforme ardite e riforme ordite*, *Riv.it. dir. proc. pen.*, 1973.

D. Pulitanò, *Diritto penale*, VIII ed. Torino, Giappichelli, p. 304 ss.

D'Alessandro, F., *Art. 40*, in *Codice penale commentato*, E. Dolcini, G. Marinucci (a cura di), III ed., 2011, p. 411 ss.

D'Alessandro, F., *Le frequenze medio-basse e il nesso causale tra omissione ed evento*, in *Cass. pen.* 2007, p. 4828.

D'Alessandro, F., *Rischio e responsabilità amministrativa degli enti*, in *Il diritto penale di fronte alle sfide della "società del rischio"*, cit., pp. 90-111.

D'Alessandro, F., voce *Precauzione (principio di)*, *Enc. dir., I Tematici, Il reato colposo*, Milano, Giuffrè, 2021, p. 975 ss.

De Amicis, G., *Il principio di legalità nella giurisprudenza delle Corti europee*, in *Quaderni europei*, ottobre 2009, n. 14.

De Francesco, G., *Dinamiche del rischio e modelli d'incriminazione nel campo della circolazione di prodotti alimentari*, in *Per uno studio interdisciplinare su agricoltura e alimentazione*, M. Goldoni, E. Sirsi (a cura di), Milano, Giuffrè, 2010, pp. 298 ss.

De Francesco, G., *L'imputazione del reato e i tormenti del penalista*, in *Scritti per Federico Stella*, cit., pp. 513 ss.

De Francesco, V., *Sulla misura soggettiva della colpa*, in *Studi Urbinati, A, Scienze giuridiche, politiche ed economiche*, 30., 1978 pp. 273-343.

Demuro, G.P., *Voce Homo eiusdem professionis et condicionis (profili storici)*, in *Ec. Dir., I Tematici, Il reato colposo*, Milano, Giuffrè, 2001, p. 607 ss.

De Vero, G., *Il nesso causale e il diritto penale del rischio*, in *Riv. it. dir. proc. pen.*, 2, 2016, p. 670 ss.

Defrance, G., *L'application de l'article L. 121-3 du code de la route*, *Jur. Auto.*, nov. 2008, p. 635.

Defrance, G., *Le propriétaire du véhicule redevable pénalement*, *Jur. Automobile*, n. 745, oct. 2003, pp. 484-487.

Delitala, G., *Il "fatto" nella teoria generale del reato*, Padova, Cedam, 1930

Desessard, L., *Commentaires au Code de la route, art. 121-1*, Dalloz, 17 ed.

Desessard, L., *Le traitement simplifié des infractions à la circulation routière*, in *Un droit pénal postmoderne ?*, Paris, Presses Universitaires de France, pp. 298 ss.

Desportes, F., Le Gunehec, F., *Droit pénal général*, 6ed., Paris, Economica.

Detraz, S., *De la mauvaise conduite des véhicules autonomes en droit pénal*, Recueil Dalloz, 2021, p. 1039.

Di Florio, M., *Causalità e colpa medica in rapporto all'emergenza sanitaria*, *Leg. pen.*, 17.09.2021, p. 4.

Di Florio, M., *Il diritto penale che verrà. Brevi considerazioni sul possibile impiego dell'IA per prevenire il rischio di disastri colposi*, in *Arch. Pen.*, 2, 2021.

Di Giovine, O., *Ancora sull'infezione da sars-cov-2: omicidio e lesioni personali tra incertezze scientifiche, sovradeterminazioni causali e trappole cognitive*, *Leg. Pen.*, 27.01.2021, p. 5.

Di Giovine, O., *La causalità omissiva in campo medico-chirurgico al vaglio delle Sezioni Unite*, in *Foro it.*, II, 2002, pp. 608 ss.

Di Giovine, O., *La causalità tra scienza e giurisprudenza*, *Riv. it. med. Leg.*, 1/2016, p. 42.

Di Giovine, O., *Lo statuto epistemologico della causalità penale tra cause sufficienti e condizioni necessarie*, *Riv. it. dir. proc. pen.*, 2002.

Di Giovine, O., *Probabilità statistica e probabilità logica nell'accertamento del nesso di causalità*, *Cass. pen.*, 2008, p. 2157.

Di Giovine, O. *Il contributo della vittima nel delitto colposo*, Torino, Giappichelli, 2003.

Di Rosa, G., *Quali regole per i sistemi automatizzati "intelligenti"?*, in *Riv. dir. civ.*, sett-ott. 2021, p. 853.

Di Rosa, G., *Robot e responsabilità per danni*, in *L'algoritmo pensante*, Il Pozzo di Giacobbe, Trapani, 2020, p. 85 ss.

Domingos, P., *L'algoritmo definitivo*, Torino, Bollati Boringhieri, 2016.

M. Donini, *Teoria del reato*, Padova, Cedam, 1996.

Donini, M., *La causalità omissiva e l'imputazione "per l'aumento del rischio"*, in *Riv. it. dir. proc. pen.*, 1999.

Donini, M., *Il garantismo della condicio sine qua non e il prezzo del suo abbandono*, in *Riv. it. dir. proc. pen.*, 2011, p. 504 ss.

Donini, *L'elemento soggettivo della colpa. Garanzie e sistematica*, in *Riv. it. dir. proc. pen.*, 2013, pp. 124-156.

Dovere, S., *Giurisprudenza della Corte suprema sulla colpa*, in *Enc Dir., I Tematici, Il reato colposo*, Milano, Giuffrè, p. 579 ss.

Dray, W. H., *Filosofia e conoscenza storica*, Bologna, Il Mulino, 1973.

Dreyer, E., *Droit pénal général*, 5 ed., Paris, Lexis Nexis, 2019.

Dreyer, E., *La causalité directe de l'infraction*, *Dr. Pèn.*, 2007, n° 6, Juin 2007, étude 9.

Dreyer, E., *Droit pénal et principe de précaution*, in *Recueil Dalloz*, 1 oct. 2015, n. 33, pp. 1912-1918.

Edmonds, D., *Uccideresti l'uomo grasso? Il dilemma etico del male minore*, Milano, Raffaello Cortina Editore, 2014.

Egelmores, R. S., *Artificial Intelligence and Knowledge Based Systems: origins, methods and opportunities for NDE*, <https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1121&context=qnde>.

Engisch, K., *Die Kausalität als Merkmal der strafrechtlichen Tatbestände*, Tübingen, Mohr, 1931.

Eschenbach, W. J., *Transparency and the Black Box Problem: Why We Do Not Trust AI*, *Phil. and Tech.*, 1. Sept. 2021.

Eusebi, L., *Appunti sul confine tra dolo e colpa nella teoria del reato*, in *Riv. it. dir. proc. pen.*, 2000, p. 1057.

Eusebi, art. 42, in *Commentario breve al codice penale*, a cura di G. Forti, S. Seminara, G. Zuccalà, VI ed., Milano, Wolters Kluwer, 2017, p. 173 ss.

Eykholt, K. et al., *Robust Physical-World Attacks on Deep Learning Visual Classification*, 2018, *arXiv.it*

Fiandaca, G., Musco, E., *Diritto Penale, P.te generale*, VIII ed., Bologna, Zanichelli.

Fiandaca, G., Musco, E., *Diritto penale, parte generale*, VI ed., Bologna, Zanichelli.

Fiandaca, G., voce *Causalità*, in *Dig. Disc. Pen.*, II, Torino, UTET, 1988.

Finocchiaro, G., *Riflessioni su diritto e tecnica*, in *Dir. dell'informazione e dell'informatica*, fasc. 4-5, 2012, p. 831 ss.

Fioriglio, G., *La "dittatura" dell'algoritmo: i motori di ricerca web e neutralità della indicizzazione. Profili informatico-giuridici*, in *Bocconi Legal Papers*, 5, 2015, p. 113-140. *HeinOnline*.

Floridi, L., *La quarta rivoluzione. Come l'infosfera sta trasformando il modo*, Raffaello Cortina Editore, Milano, 2017.

Foot, P.R., *The Problem of Abortion and the Doctrine of the double Effect*, *Oxford Review*, V, 1967, pp. 5-15.

Fornasari, G., *Il principio di inesigibilità nel diritto penale*, Padova, CEDAM, 1990.

Forti, G. *Colpa ed evento nel diritto penale*, Milano, Giuffrè, 1990.

Gadamer, H. D., *Verità e metodo*, Bompiani, Milano, 1983.

Galetta, D.U., Corvalán, J.G., *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *Federalismi.it*, n. 3/2019.

Gargani, A., *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato di fronte alle esigenze di controllo penale delle nuove fenomenologie di rischio*, in *LP*, 2, 2011, p. 400.

Gauvin, F., *obs. Cass. Crim., 18 juin 2002, n. 01-86.539, Recueil Dalloz*, p. 240.

Gazzoni, F., *Manuale di diritto privato, XVI ED.*, Napoli, Edizioni Scientifiche Italiane, 2013, p. 734.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Dumé, H., Crawford, K., *Datasheets for Datasets*, <https://arxiv.org/abs/1803.09010>.

Gialuz, M., *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra stati uniti ed europa*, *Dir. Pen. Cont.*, [www.penalecontemporaneo.it](http://www.penalecontemporaneo.it).

Giuffrida, *Liability for AI Decision-Making: Some Legal and Ethical Considerations*, 88 *FORDHAM L. REV.* 439 (2019)

Giunta, F., *Illiceità e colpevolezza nella responsabilità colposa. I - La Fattispecie*; Padova, Cedam, 1993.

Giunta, F., *La normatività della colpa penale. Lineamenti di una teorica*, in *Riv. it. dir. proc. pen.*, 1999, p. 86-115.

Giunta, F., *Il diritto penale e le suggestioni del principio di precauzione*, in *Criminalia*, 2006, p. 227 ss.

Giunta, F., *Culpa, culpae*, in *disCrimen*, 4.06.2019

Gless, S., Silverman, E., Weigend, T., *If Robots Cause Harm, Who Is to Blame? Self-Driving Cars and Criminal Liability*, (January, 29, 2016), SSRN: <http://dx.doi.org/10.2139/ssrn.2724592>

- Goodfellow, I., Bengio, Y., Courville, A., *Deep learning*, MIT Press, 2016.
- Grasso, G., Giuffrida, F., *Gli effetti della giurisprudenza della Corte Edu sull'ordinamento italiano: prospettive di diritto penale sostanziale, L'indice penale*, Gennaio-aprile 2020.
- Grasso, G., *Il reato omissivo improprio*, Milano, Giuffrè, 1983.
- Grasso, G., *Il trattato di Lisbona e le nuove competenze penali dell'Unione europea*, in *Studi in onore di Mario Romano, IV*, Napoli, Jovene, 2011, pp. 2307 ss.
- Greenberg, R. S., *Medical Epidemiology. Population Health and Effective Health Care, 5 ed.*, McGraw-Hill Education, 2015.
- Grosso, *Difesa legittima e stato di necessità*, Milano, Giuffrè, 1964.
- Grotto, M., *Principio di colpevolezza, rimproverabilità soggettiva e colpa specifica*, Torino, Giappichelli, 2012.
- Grotto, M., voce *Scusanti della colpa*, in *Enc. dir., I Tematici, Reato colposo*, Giuffrè, Milano, 2021, pp 1153 ss.
- Hadri, S., *Pleins phares sur les évolutions du cadre légal et réglementaire des véhicules autonomes, Petites affiches*, n. 3, 3 janv. 2019.
- Hallevy, G., *Liability for Crimes and Artificial Intelligence*, Springer, 2014.
- Hallevy, G., *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, Berlino, 2015.
- Hallevy, G., *The Basic Models of Criminal Liability of AI Systems and Outer Circles*, 11 giugno 2019, *ssrn.com*.
- Hallevy, G., *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, *Akron Intellectual Property Journal*, Vol. 4 : Iss. 2 , Article 1, 2010.
- Hallevy, G., *When Robots Kill: Artificial Intelligence under Criminal Law*, Northeastern University Press, 2013.
- Halpérin, J.L., *Introduction au droit*, Paris, Dalloz, 2019.
- Hebb, D., *L'organizzazione del comportamento*, Franco Angeli, 1975.
- Hempel, C. G., Oppenheim, P., *Studies in the logic of explanation*, in *Philosophy of Science, vol. 15*, 1948, pp. 135-175.
- Hempel, C.G., *Aspetti della spiegazione scientifica*, Milano, Il Saggiatore, 1987.

Hornbek, M., *AI has just revealed there are over 1.8 billion trees in the Sahara Desert* in <https://www.weforum.org/agenda/2020/10/ai-billion-trees-sahara-desert-technology/>.

Hubbard, F. P., “*Sophisticated Robots*”: *Balancing Liability, Regulation, and Innovation*, 66 FLA. L. REV. (2014), 1803 ss.

Humphreys, P., *The philosophical novelty of computer simulation methods*, *Synthese*, 2009, 169(3), pp. 615–626.

Hutson, M., *Self-taught Artificial Intelligence Beats Doctors at Predicting Heart Attacks*, *Science*, 2017.

Irti, Severino, *Le domande del giurista e le risposte del filosofo (un dialogo su diritto e tecnica)*, in *Contratto e impresa*, 2006, p. 665.

Italiano, G. F., *Intelligenza artificiale: passato presente, futuro*, in *Intelligenza artificiale, protezione dei dati personali e regolazione*, F. Pizzetti (a cura di), Giappichelli, Torino, 2018, p. 209.

Josseume, R., *La force majeure en droit routier: un mirage législatif?*, *Gaz. Pal., avr. 2018, n. 16*, pp. 19-21.

Joynt, C.B., Rescher, N., *The problem of Uniqueness in History*, in *History and Theory*, I, 1961, n. 2, p. 154.

Kaplan, J., *Intelligenza artificiale. Guida al futuro prossimo*, LUISS University Press, 2017.

Kaufmann, A., *Tatbestandmässigkeit und Verursachung im Contenganverfahren*, in *JZ*, 1971, p. 569.

Krohn, J., Grant, B., Aglaé, B., *Deep Learning Illustrated*, Addison-Wesley Professional, 2019.

Kroll, J. A., *The fallacy of inscrutability*, *Phil. Trans. R. Soc. A* 376.

Kurzweil, R., *The age of intelligent machines*, MIT Press, 1990.

Kurzweil, R., *La singolarità è vicina*, Apogeo Education - Maggioli Editore, 2008.

F. Lagioia, G. Sartor, *AI Systems Under Criminal Law. A Legal Analysis and A Regulatory Perspective*, in *Phil.&Techn.*, 2019.

Lagioia, F., Sartor, G., *L'intelligenza artificiale per i diritti dei cittadini: il progetto Claudette*, in *Ragion Pratica*, I, giugno 2020, p. 91.

Leiman, T., *Law and tech collide: foreseeability, reasonableness and advanced driver assistance systems*, *Policy and Society*, (2021), 40:2, 250-271.

Lerman, J., *Big Data and its Exclusions*, 66, *Stanford Law Review online*, 2013, pp. 55 ss.

Levade, A., *Conseil constitutionnel et ordonnances : l'invraisemblable revirement !*, *La Semaine juridique. Édition générale*, 29 juin 2020, n° 26, p. 1185 – 1186

Levade, A., *Ordonnance et loi d'habilitation, le Conseil constitutionnel étend son empire*, *La semaine juridique - édition générale - n° 30-35 - 27 juillet 2020*.

Lipton, Z. C., *The mythos of model interpretability*, *ArXiv*, 2016, p. 1606.

Listz, F., *Leherbuch des deutschen Strafrechts*, De Gruyter, 1894 pp. 108 ss.

Lotti, M., *Mesotelioma: scienza e congetture nel processo penale*, in *Processo alla scienza*, R. Borsari (a cura di), Padova, Padova University Press, 2016.

Luchaire, F., *Deux décisions du Conseil constitutionnel appelées à faire jurisprudence ?*, *Rev. Dr. Publ.*, 5, 1999, p. 1283-1292.

Luden, *Abhandlungen aus dem gemeinen deutschen Strafrecht*, vol. I, 1836.

Lunghini, G., Paris, L., *art. 45*, in *Codice penale commentato*, E. Dolcini, G. Marinucci, (diretto da) E. Dolcini, G. L. Gatta, Milano, Wolters Kluwer, V ed., 2021, p. 726-727

Ludwiczak, F., *Responsabilité pécuniaire de l'art. L. 121-3 du code de la route : de l'incertitude à l'imprecision*, *Rev. Pénitentiaire*, 2013, p. 547-554.

Maiwald, M., *Causalità e diritto penale*, Milano, Giuffrè, 1999.

Manes, V., *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *DisCrimen*, 15.5.2020.

Manna, A., *I rapporti tra sapere scientifico e sapere giudiziario*, in *Cass. pen.*, 2009, p. 3636.

Mantovani, F., *voce Colpa*, in *Dig. Disc. Pen.*, 1988, vol. II, Padova, Cedam, pp. 299-314.

Mantovani, F., *Diritto penale, Parte generale, X ed.*, Milano, Wolters Kluwer, CEDAM, 2019.

Mantovani, M., *Il principio di affidamento nella teoria del reato colposo*, Milano, Giuffrè, 1997.

Marchant, G.E., Lindor, R.A., *The Coming Collision Between Autonomous Vehicles and the Liability System*, 52 *Santa Clara l. Rev.* 1321, 1334 (2012).

Marconi, D., *Linguaggio e scienza cognitiva*, in *Enc. Treccani, XXI Secolo, Norme e Idee*, 2009, pp. 201-210.

Marini, G., voce *Colpevolezza*, in *Dig. Disc. Pen.*, 1988, vol. II, Padova, Cedam, pp. 314-322.

Marino, G., *La Germania dice sì alla guida autonoma: è la prima in Europa*, *www.repubblica.it*, 31 maggio 2021.

Marinucci, G., *La colpa per inosservanza di leggi*, Milano, Giuffrè, 1965.

Marinucci, G., *Il reato come azione*, Milano, Giuffrè, 1971.

Marinucci, G., *Innovazioni tecnologiche e scoperte scientifiche: costi e tempi di adeguamento delle regole di diligenza*, in *Riv. it. dir. proc. pen.*, 2005, I, pp. 30-60.

Marinucci, G., *Causalità reale e causalità ipotetica nell'omissione impropria*, in *Riv. it. dir. proc. pen.*, 2009, pp. 523 ss.

Marmo, R., *Algoritmi per l'intelligenza artificiale*, Hoepli, Milano, 2020.

Marzano, G. M., *Brevi note sulla prova della causalità nel contagio da Covid-19*, *Cass. pen.*, 2020, pp. 3114-15.

Masera, L., *Accertamento alternativo e evidenza epidemiologica nel diritto penale. Gestione del dubbio e profili causali*, Milano, Giuffrè, 2007.

Masera, L., *Ancora sulla qualificazione penalistica dell'evidenza epidemiologica perché anche nella società del rischio è legittimo il ricorso al diritto penale d'evento*, in *Il diritto penale di fronte alle sfide della "società del rischio"*, G. De Francesco, G. Morgante (a cura di), Torino, Giappichelli, pp. 389 ss.

Masera, L., *Danni da amianto e diritto penale*, *www.penalecontemporaneo.it*.

Masera, L., *Evidenza epidemiologica di un aumento di mortalità e responsabilità penale*, *Riv. trim. dir. pen. cont.*, 3-4, 2014, pp. 343-365.

Masera, L., *Il modello causale delle Sezioni unite e la causalità omissiva*, in *Dir. pen. proc.*, 2006, 4, pp. 493 ss.

Mathieu, B., *La sécurité routière fait progresser le droit constitutionnel pénal*, *Recueil Dalloz*, 15 juill 1999, n. 27.

Maugeri, A.M., *L'uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra evidence based practices e tutela dei diritti fondamentali*, in *Arch. Pen.*, 1, 2021, pp. 1-37.

Mayaud, Y., *Droit pénal général*, 6 ed., Paris, Presses Universitaires de France (PUF), 2018.

Mayaud, Y., *Entre le dit et le non-dit, ou les leçons de droit pénal du Conseil constitutionnel*, *Recueil Dalloz*, 1999, p. 589.

- Mayaud, Y., *obs. Cass. Crim. 25 sept. 2001, Bull. crim. n. 188, RSC 2002*, p. 101.
- Mayaud, Y., *obs. Cass. Crim. 29 oct. 2002, Bull. crim. n. 196, RSC 2003*, p. 330.
- Mayaud, Y., *obs. Cass. crim., 25 sept. 2001, Bull. crim. 2001, n° 188 ; RSC 2002*, p. 101.
- Mayaud, Y., RSC 2003, pp. 330 ss.
- Mayaud, Y., *obs. Cass. Crim. 25 sept. 2001, Bull. crim. n. 188*, in *Revue de science criminelle et de droit pénal comparé (RSC)*, 2002, p. 101
- Mayaud, Y., *obs. Cass. crim. 5 oct. 2004, Bull. Crim. n. 230; RSC 2005*, p. 71.
- McCarthy, J., Minsky, M. L., Rochester, N., Shannon, C.E., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955.
- McCarthy, J., *Programs with Common Sense*, 1958, <http://jmc.stanford.edu/articles/mcc59/mcc59.pdf>.
- Mehta, P., Bukov, M., Wang C.H., et al., *A hig-bias, low-variance introduction to Machine Learning for physicists*, in *Physics Reports 810* (2019) 1–124.
- Mezzetti, E., *Colpa per assunzione*, in *Scritti in memoria di Giuliano Marini*, Jovene, Napoli, 2010, pp. 513 ss.
- Militello, V., *Rischio e responsabilità penale*, Milano, Giuffré, 1988.
- Minsky, M., Papert, S., *Perceptrons. An Introduction to Computational Geometry*, Mit Press, 1969.
- Miotto, R. et al. *Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records*, *Sci. Rep.* 6, 26094, (2016).
- Monot-Fouletier, M., *Véhicule autonome: vers une autonomie du régime de responsabilité applicable ?*, Recueil Dalloz, 2018, p. 129.
- Monot-Fouletier, M., *Véhicule autonome: vers une autonomie du régime de responsabilité applicable ?*, Recueil Dalloz, 2018, p. 129.
- Montavon, G., Samek, W., Müller, K.R., *Methods for interpreting and understanding deep neural networks*, *Digital signal processing*, 73, 2018, pp. 1-15.
- Murphy, K. P., *Machine learning. A probabilistic perspective*, The MIT Press, Massachusetts, 2012.
- Murri, Augusto, *Quattro lezioni e una perizia*, Bologna, 1972.
- Nagel, E., *La struttura della scienza*, Milano, Feltrinelli, 1968.

Nilsson, N. J., *The Quest for Artificial Intelligence. A History of Ideas and Achievements*, Cambridge University Press, 2010.

Nobili, *Nuove polemiche sulle cosiddette "massime di esperienza"*, *Riv. it. dir. proc. pen.*, 1969, p. 123 ss.

Orsina, A., *Rischio da incertezza scientifica e modelli di tutela penale*, Giappichelli, Torino, 2015.

Pagallo, U., *The Laws of Robots. Crimes, Contracts, and Torts*, Springer, Dordrecht, 2013, p. 10.

Palazzo, F., *Morti da amianto e colpa penale*, *Dir. pen. proc.*, 2011, pp. 186 ss.

Paliero, C. E., *Causalità e probabilità tra diritto penale e medicina legale*, *Riv. It. med. Leg.*, 4, 2015, p. 1515.

Paliero, C.E., *La causalità dell'omissione: formule concettuali e paradigmi prasseologici*, in *Riv. it. med. Leg.*, XIV, 1992, p. 842.

Paliero, C.E., *Responsabilità per danno da prodotto tra persona fisica ed ente collettivo*, in *Il diritto penale di fronte alle sfide della "società del rischio"*, cit., p. 291.

Parisi, D., voce *Intelligenza artificiale*, in *Enciclopedia delle scienze sociali*, Treccani, 1996, in [www.treccani.it](http://www.treccani.it).

Pasquinelli, A., *Nuovi principi di epistemologia*, Milano, Feltrinelli, 1974, p. 101.

Pecoraro Albani, A., voce *Caso fortuito (dir. pen.)*, in *Enc. dir.*, VI, Milano, Giuffrè, 1960, p. 391.

Piemontese, C., *Il principio dell'"oltre il ragionevole dubbio", tra accertamento processuale e ricostruzione dei presupposti della responsabilità penale*, *Dir. pen. proc.*, 2004, 6, pp. 757 ss.

Piergallini, C., *Danno da prodotto e responsabilità penale*, Milano, Giuffrè, 2004.

Piergallini, C., voce *Colpa (diritto penale)*, *Enc. dir. – Annali X*, 2017, pp. 222 -265.

Piergallini, C., *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, in *Riv. it. dir. proc. pen.*, 4, 2020, pp. 1745-1774.

Pin, X., *Droit pénal général*, 12 ed., Paris, Dalloz, 2020, p. 198.

Pisani, N., *La "colpa per assunzione" nel diritto penale del lavoro*, Jovene, Napoli, 2012.

Pisani, N., voce *Colpa per assunzione*, in *Ec. dir. I Tematici, Il reato colposo*, Milano, iuffrè, 2021, pp. 233 ss.

Piva, D., *Spunti per una riscoperta della colpa per assunzione*, in *DisCrimen*, 9.9.2020.

Planck, M., *La conoscenza del mondo fisico*, Torino, 1949.

Popper, Karl, *Logica della scoperta scientifica*, Torino, Einaudi, 1970.

Prince, A., Schwarcz, D.B., *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, in *Iowa Law Review*, 5 agosto 2019.

Prothais, A., *obs. Crim. 18 juin 2003, Bull. crim. n° 127 ; Recueil Dalloz, 2005. Jur.* 195.

Pulitanò, D., *Igiene e sicurezza del lavoro (tutela penale)* in *Dig. Disc. Pen. – Aggiornato 2000*, p. 396 ss.

Pulitanò, D., *Diritto Penale, VIII ed.*, Torino, Giappichelli, 2019.

Pulitanò, D., *Gestione del rischio da esposizioni professionali*, *Cass. pen.*, 2, 2006, p. 789.

Quattrococo, S., *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale “predittiva”*, *Cass. pen.*, 4, 2019, pp. 1748 ss.

Radbruch, G., *Zur Systematik der Verbrechenslehre*, in *Festgabe für Frank*, 1930, ristampa anast. 1969, vol. I.

Resta, G., *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, in *Politica del diritto*, 50, 2019, n. 2, pp. 199-236.

Ribeiro, M. T., Singh, S., Guestrin, C., *“Why Should I Trust You?”: Explaining the Predictions of Any Classifier*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, pp. 1135–1144. ACM, 2016.

Robert, J.H., *obs. Cass. crim.*, 31 janv. 2007, n. 07-81.365, *Dr. Pén.* 2007, comm 138.

Rodotà, S., *Il modno nella rete. Quali i diritti, quali i vincoli*, Laterza, Roma-Bari, VI ed., 2019.

Romano, M., *Commentario sistematico del codice penale, art. 40*, Milano, Giuffrè, 2004, p. 373.

Romano, M., *Nesso causale e leggi scientifiche in diritto penale*, in *Scritti per Federico Stella*, Napoli, Jovene, 2007.

Rosenblatt, F., *The Perceptron: a probabilistic model for information storage and organization in the brain*, in *Psychological Review*, Vol. 65, n. 6, 1958.

Rudin, C., *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*, in *Nature Machine Intelligence*, 1(5), 206–215, May 2019.

Ruffolo, U., Al Mureden, E., *Intelligenza artificiale e diritto. Autonomous vehicles e responsabilità nel nostro sistema e in quello statunitense*, in *Giur. It.*, 2019, 7, p. 1704 ss.

Ruffolo, U., *Intelligenza artificiale, machine learning e responsabilità da algoritmo*, in *Giur. it.*, luglio 2019, p. 1693.

Ruffolo, U., *Per i fondamenti di un diritto della robotica self-learning; dalla machinery produttiva all'auto driverless: verso una "responsabilità da algoritmo"?*, in *Intelligenza artificiale e responsabilità*, a cura di U. Ruffolo, Milano, Giuffrè, 2017, p. 3.

Ruga Riva, C., *Principio di precauzione e diritto penale. Genesi e contenuto della colpa in contesti di incertezza scientifica*. In E. Dolcini, & C.E. Paliero (a cura di), *Studi in onore di Giorgio Marinucci* (pp. 1743-1777), Milano, Giuffrè, 2006.

Ruiz de Villa Robert, A., Puertas Prats, E., *I big data, il machine learning e la statistica*, volume II della collana *La matematica che trasforma il mondo*, RBA editore, 2020.

Russel, S. J., Norvig, P., *Intelligenza artificiale, un approccio moderno*, Utet, Torino, 2002.

Salvadori, I., *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in *Riv. it. dir. proc. pen.*, 1, 2021, pp. 83- 118.

Santosuosso, A., Boscarato, C., Caroleo, F., *Robot e diritto: una prima ricognizione*, in *Giur civ.*, 2012, 7-8, 20494 ss.

Savignac, J., *Tools to Identify and Assess the Risk of Offending Among Youth*, Published by *National Crime Prevention Centre (NCPC)*, 2010.

Searle, J. R., *Is the Brain's Mind a Computer Program?*, in *Scientific American*, january 1990, pp. 26-32.

Selbst, A. D., *Negligence and AI's Human Users*, in *100 Boston University Law Review*, 2020, 1315 ss.

Sjöström, H., Nilsson, R., *Il talidomide e il potere dell'industria farmaceutica*, Milano, Feltrinelli, 1973.

Sofge, E., *Artificial Intelligence Will Not Obliterate Humanity*, *Popular Sci.*, Mar. 19, 2015, p. 36 ss.

Somalico, M., Amigoni, F., Schiaffonati, V., *La grande scienza. Intelligenza artificiale*, in *Storia della scienza, 2003, Enciclopedia Treccani online, www.treccani.it*.

Smith, B. W., *Automated Driving and Product Liability*, *Michigan State Law Review* 2017, no. 1 (2017), pp. 1-74.

Stella, F., *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, 3 ed. Milano, Giuffrè, 2003.

Stella, F., *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, Milano, Giuffrè, 2003.

Stella, F., *Il giudice corpuscoliano. La cultura delle prove*, Milano, Giuffrè, 2005.

Stella, F., *La nozione penalmente rilevante di causa: la condizione necessaria*, in *Riv. it. dir. proc. pen.*, 1988, pp. 1249 ss.

Stella, F., *Leggi scientifiche e spiegazione causale nel diritto penale*, Milano, Giuffrè, ristampa.

Stella, F., *Verità, scienza e giustizia: le frequenze medio-basse nella successione di eventi*, *Riv. it. dir. proc. pen.*, 2002, p. 1236.

Stortoni, L., *Angoscia tecnologica ed esorcismo penale*, in *Riv. it. dir. proc. pen.*, 2004.

Summerer, K., *Causalità ed evitabilità. Formula della condicio sine qua non e rilevanza dei decorsi causali ipotetici nel diritto penale*, Pisa, Edizioni ETS, 2013.

Summerer, K., voce *Evitabilità dell'evento e comportamento alternativo lecito*, entrambi in *Enc. Dir., I Tematici, Il reato colposo*, Milano, Giuffrè, 2021, p. 489 ss.

Tafari, D., *Sulla moralità artificiale. Le decisioni delle macchine tra etica e diritto*, in *Riv. di filosofia*, 1, aprile 2020, pp. 83 ss..

Taruffo, M., *La prova del nesso causale*, in C. De Maglie, S. Seminara (a cura di), *Scienza e causalità*, Padova, CEDAM, 2006, p. 90 ss.

Teresi, L., *Véhicule à délégation de conduite et risque automobile : une lecture juridique*, 2020, [actu-juridique.fr](http://actu-juridique.fr).

Terrett, A., *Neural Networks - Towards Predictive Law Machines*, 3 *Int'l J.L. & info. Tech.*, 94, 1995.

Thiene, A., *art. 2051 cc.*, in G. Cian, A. Trabucchi, *Commentario breve al codice civile*, XIV ed., CEDAM, Wolters Kluwer, 2020, p. 2231.

Thomson, J.J., *Killing, letting die and the trolley problem*, in *The Monist*, LIX, 1976, pp. 204-217

Tola, S., Miron, M., Gómez, E., Castillo, C., *Why machine learning may lead to unfairness: evidence from risk assessment for Juvenile Justice in Catalonia*, Best Paper Award, International Conference on AI and Law, 2019.

Tomsett, R., Braines, D., Harborne, D., Preece, A., Chakraborty, S., *Interpretable to whom? A role based model for analyzing interpretable machine learning systems*, 2018, *arXiv.org*, p. 9.

Torchiani, G., *Auto a guida autonoma : cosa sono e come funzionano*, 18 maggio 2021, *www.ai4business.it*.

Torrente, A., Schlesinger, P., *Manuale di diritto privato, XIX ed.*, Milano, Giuffrè, 2009, p. 866.

Traversa, G., *Evidenza epidemiologica e decisioni nelle aule di tribunale*, in *Scienza, diritto e processo penale nell'era del rischio*, A. Amato, G. Flora, C. Valbonesi (a cura di), Torino, Giappichelli, 2019, p. 259.

Trezza, R., *L'Intelligenza Artificiale come ausilio alla standardizzazione del modello 231: vantaggi "possibili" e rischi "celati"*, in *Giur. pen.*, 2021/1-Bis.

Turing, A., *Computing, Machinery and Intelligence*, in *Mind*, LIX (236), 1950, pp. 433-460.

Ubertis, G., *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Sistema penale* 2020.

Venanzoni, A., *La valle del perturbante: il costituzionalismo alla prova delle intelligenze artificiali e della robotica*, in *Politica del diritto*, 2/2019.

Veneziani, P., *Il nesso tra omissione ed evento nel settore medico: struttura sostanziale ed accertamento processuale*, in *Studi in onore di Giorgio Marinucci*, E. Dolcini, C.E. Paliero (a cura di), 2006, p. 1970.

Veneziani, P., *Regole cautelari proprie e improprie nella prospettiva delle fattispecie colpose causalmente orientate*, Padova, Cedam, 2003

Vermelle, *obs. Cass. Crim. 19 oct. 2004, n. 04-82.485, RSC 2005*, p. 291.

Véron, M., *La certitude du lien de causalité, Droit pénal n° 6, Juin 2008, comm. 82*.

Véron, M., *obs Cass. Crim, 26 juin 2001, Dr. Pén. 2001, comm. 124*.

Véron, M., *obs. Cass. Crim. 22 mars 2005, Dr. pénal 2005. Comm. 103*.

Véron., M., *obs. Cass. crim. 15 oct. 2002, Bull. Crim. 2002, n. 186; Dr. Pén. 2003, comm. 4*.

Verpeaux, M., *Actes administratifs - Ordonnances = dispositions législatives. A propos de la décision du Conseil constitutionnel du 28 mai 2020* », *La Semaine juridique. Administrations et collectivités territoriales*, 22 juin 2020, n° 25, p. 4.

Viganò, F., *Stato di necessità e conflitti di doveri. Contributo alla teoria delle cause di giustificazione e delle scusanti*, Milano, Giuffrè, 2000.

Viganò, F., *Il principio di prevedibilità della decisione giudiziale in materia penale*, *Dir. pen. cont.*, 19 dicembre 2016.

Viganò, F., *Il rapporto di causalità nella giurisprudenza penale a dieci anni dalla sentenza Franzese*, *Riv. Trim. dir. pen. cont.*, 3, 2013, p. 380 ss.

Villanueva, A., *The Legal Battle with the Future of Autonomous Surgical Robotics*, *17 IND. HEALTH L. REV.* 367 (2020).

Vingiano-Viricel, I., *Véhicule autonome: qui est responsable ?*, Paris, Lexis Nexis, 2019.

Wallach, W., Allen, C., *Moral Machines: Teaching Robots Right from Wrong*, Oxford, Oxford University Press, 2009.

Zednik, C., Boelsen, H., *Preface: Overcoming Opacity in Machine Learning*, in *AISB 2021 Symposium Proceedings*, cit., p. 1.

Zednik, C., *Solving the Black Box Problem: a normative framework for explainable AI*, *Phil.Tech.*, 2021, p. 267.

Zheng, A., *Evaluating Machine Learning Models*, O'REILLY, 2015.

Zirulia, S., *Esposizione a sostanze tossiche e responsabilità penale*, Milano, Giuffrè, 2018.

Zuddas, P., *Intelligenza artificiale e discriminazioni*, *Liber amicorum per Pasquale Costanzo*, in *Consulta Online*, 16 marzo 2020, p. 5.