

**LUISS**



Dipartimento  
di Giurisprudenza

Cattedra di Metodologia della Scienza Giuridica

**CHI RISPONDE DEL ROBOT:  
IL PROBLEMA DELLA RESPONSABILITÀ  
PENALE PER I DANNI PROVOCATI DALLE  
SELF DRIVING CAR**

**Prof. Antonio Punzi**

---

**RELATORE**

**Prof. Gianluigi Ciacci**

---

**CORRELATORE**

**Bianca Napolitano - Matr. 139823**

---

**CANDIDATA**

Anno Accademico 2019/2020

*Ad Alto*

*Tutto è ignoto: un enigma, un inesplicabile mistero. Dubbio, incertezza, sospensione del giudizio appaiono l'unico risultato della nostra più accurata indagine in proposito. Ma tale è la fragilità della ragione umana, e tale il contagio irresistibile delle opinioni, che non è facile tener fede neppure a questa posizione scettica, se non guardando più lontano e opponendo superstizione a superstizione, in singolar tenzone; intanto, mentre infuria il duello, ripariamoci felicemente nelle regioni della filosofia, oscure ma tranquille.*

D.Hume

# INDICE

<b>INTRODUZIONE</b>	5
<b>Scopo obiettivi e limiti dell'indagine</b>	
<b>CAP1. CONTESTO DI RIFERIMENTO</b>	10
<b>1. Importanza dell'analisi del contesto di riferimento</b>	11
Dal cibernetica di Platone al giurista della quarta rivoluzione industriale: perché conoscere il contesto	
<b>2. Intelligenza artificiale</b>	11
2.1. Definizione tecnica e considerazioni filosofiche: <i>A.I. matrimonio o divorzio tra intelligenza e capacità di agire?</i>	
2.2. Modalità di funzionamento	
2.3. Self learning e introduzione delle questioni giuridiche e morali	
2.4. La quarta rivoluzione industriale	
2.5. Le diverse applicazioni dell'A.I.	
<b>3. Auto a guida autonoma</b>	31
3.1 AV come simbolo della quarta rivoluzione	
3.2 Quali vantaggi	
3.3 Ostacoli e problemi aperti	
3.4 Stato di evoluzione attuale: i soggetti coinvolti.	
<b>CAP 2. CAUSE DA CUI SORGE LA RESPONSABILITÀ</b>	43
<b>1. Introduzione</b>	44
1.1 Il nucleo della questione: il responsibility gap per i danni da prodotti A.I.	
1.2 Distinzione tra i tipi di danno.	
<b>2. Danni derivanti da malfunzionamento</b>	47
2.1. La riduzione delle aspettative di infallibilità a seguito dell'incidente di Uber	
2.2. La società del rischio: introduzione al concetto di rischio	
2.3. Tipologie di danno da prodotto	
2.4. Danno da machine learning: il problema del responsibility gap	
2.5. Ethic by design come possibile soluzione	
<b>3. Danni derivanti da scelte: la questione etica</b>	63
3.1 il <i>Dilemma del carrello</i> : legame tra problematiche etiche e giuridiche	
3.2 Risoluzione del problema tramite teorie deontologiche e consequenzialistiche	
3.3 Società produttrici, Stato o conducenti: chi dovrebbe decidere del carrello?. (e introduzione delle relative conseguenze sul piano della responsabilità)	
3.4 Una quarta via	

**CAP 3. REGOLAMENTAZIONE DELLA RESPONSABILITÀ: STATO DELL'ARTE E PROSPETTIVE** 77

**1. Il contesto giuridico e sociale di riferimento** 78

- 1.1 Principi costituzionali in materia penale e difficoltà di ripercorrere le strade individuate dal Parlamento Europeo
- 1.2 La società delle mangrovie
- 1.3 Necessità di distinguere le cause da cui scaturisce il danno

**2. Responsabilità per i danni derivanti da malfunzionamento** 86

- 2.1. Allocazione della responsabilità penale per danni causati da vetture semi-autonome e il control dilemma
- 2.2. Allocazione della responsabilità per danni causati da vetture completamente autonome
- 2.3. Una prima proposta: Robot come nuovo soggetto di diritto
  - La proposta che chiude il cerchio della quarta rivoluzione
  - Da cosa nasce l'esigenza di responsabilizzare direttamente il robot
  - Opinioni favorevoli all'attribuzione della responsabilità direttamente al robot: la tesi di Hallevy
  - Il problema delle sanzioni per i prodotti dell'intelligenza artificiale
  - Opinioni contrarie all'attribuzione della responsabilità direttamente al robot.
- 2.4. Una seconda proposta: la responsabilità del produttore e del programmatore
  - L'estensibilità della categoria penalistica per danno da prodotto.

**3. Responsabilità per i danni provocati da scelte** 127

- 3.1 Introduzione
- 3.2 Analisi di come varia la allocazione della responsabilità in scenari di emergenza a seconda del grado di autonomia dell'agente artificiale
- 3.3 Stato di necessità e dibattito sull'operatività dello stesso con riferimento alla pre-programmazione degli AV
  - Stato di necessità come scusante
  - Stato di necessità come causa di giustificazione
  - Stato di necessità in assenza dell'attualità del pericolo
- 3.4 In che direzione dovrebbe muoversi il programmatore per essere esente da responsabilità
  - A favore di una soluzione consequenzialistica
  - Contro una soluzione consequenzialistica
  - L'esigenza di rispettare il dovere di non danneggiare e proteggere
  - L'attuale soluzione

**CONCLUSIONI** 153

## **INTRODUZIONE**

### **Scopo, obiettivi e limiti dell'indagine**

In questa analisi si vuole ragionare sui temi etico-giuridici sollevati dalle nuove tecnologie e, più nello specifico, si affronterà il problema dell'allocazione della responsabilità penale per i danni provocati dai diversi prodotti dell'intelligenza artificiale, cogliendone di conseguenza le relative implicazioni sul piano giuridico, e prima ancora etico. L'obiettivo che ci si propone di raggiungere è pertanto quello di intravedere e prevedere come, in un futuro più o meno prossimo, la questione relativa all'attribuzione della responsabilità penale per i danni provocati dai prodotti dell'intelligenza artificiale possa essere regolata. Ci si addenterà in un terreno ancora privo di una effettiva regolamentazione ma tuttavia intessuto di una fitta rete di dottrine, teorie e proposte molte delle quali indirizzate verso uno stravolgimento degli ordinari canoni del diritto penale classico e spesso tra loro contrastanti.

Al fine quindi di impostare nella maniera più corretta possibile l'analisi, sarà doveroso procedere individuandone tutti i possibili aspetti, riscontri e i relativi profili di criticità non solamente da un punto di vista prettamente giuridico ma piuttosto adottando uno sguardo e una lettura che tengano in primaria considerazione le connesse questioni etiche e morali.

Si tratta difatti di una materia in cui, data la sua particolarità e la sua delicatezza, le questioni etiche e quelle giuridiche non sembrano poter essere scisse le une dalle altre ma vadano al contrario di pari passo, intrecciandosi in uno stretto legame che ne appare impedire la risoluzione in via autonoma e separata. È per questo motivo che nel corso dell'analisi si accennerà costantemente a un doveroso parallelismo concettuale tra le diverse problematiche attinenti all'ambito giuridico e gli insegnamenti dei grandi filosofi, da quelli più antichi a quelli attuali e, per ogni ipotesi e proposta elaborata dalla più recente dottrina penalistica, si ricercherà da un lato il fondamento etico che vi risiede alla base e dall'altro le ripercussioni e le conseguenze sociali che da esse potrebbero derivare. Si ritiene tuttavia opportuno precisare fin da subito come nel corso dell'analisi si insisterà più volte sul concetto secondo cui, il fatto che il potenziale utilizzo di detti prodotti ci ponga dinnanzi a difficili quesiti etici e giuridici, non dovrebbe rappresentare un elemento tale da impedire di cogliere le opportunità economiche e sociali offerte dallo sviluppo delle nuove tecnologie, ma piuttosto sarà necessario ragionare su tali problemi e arricchire

l'attuale consapevolezza della tematica al fine di governare al meglio quel futuro da cui inevitabilmente saremo destinati ad essere travolti. Si ritiene dunque che condurre l'indagine in tale maniera, tenendo conto di ogni possibile variante, sia l'unica via per una disamina completa dell'argomento.

Si partirà dunque da un'analisi di carattere generale del contesto di riferimento all'interno del quale scaturiscono le diverse problematiche relative all'attribuzione della responsabilità. Si analizzerà il concetto di intelligenza artificiale, le relative modalità di funzionamento e le sue principali applicazioni, tra cui rientrano le automobili driverless. Al fine di affrontare nella maniera più esaustiva possibile le questioni etico giuridiche connesse alla tematica si ritiene infatti necessario conoscere la materia che si dovrà regolare e soprattutto comprendere il modo tramite cui tali prodotti sono in grado di agire e, di conseguenza, porre in essere azioni potenzialmente dannose, idonee a far sorgere problematiche connesse all'allocazione della relativa responsabilità penale.

Nel secondo capitolo si inizierà ad affrontare il problema della responsabilità penale discutendo così delle ragioni che rendono la risoluzione dello stesso particolarmente complessa. Ci si soffermerà in particolar modo sulla principale di tale problematiche, ben tratteggiata dal Parlamento Europeo nel c.d. *responsability gap*: quel vuoto di responsabilità tale da far sorgere l'interrogativo circa la possibilità di estendere le norme attualmente esistenti anche a tali tipologie di danno ovvero, al contrario, a provvedere alla creazione di disposizioni ad hoc.

Al fine di impostare l'analisi nella maniera più approfondita possibile, individuando così quale tra le due strade alternative individuate in sede istituzionale appaia opportuno percorrere si è ritenuto doveroso distinguere a monte, in due diversi filoni, le cause da cui il danno può essere generato: *danni derivanti da malfunzionamento* e *danni derivanti da scelte*. Il primo di essi fa riferimento a un'evoluzione anomala e non prevista dell'algoritmo e, in tali casi, il problema della responsabilità verrà impostato ipotizzando un'estensione o una modifica delle attuali disposizioni relative alla responsabilità per danno da prodotto difettoso. Si è dunque proceduto a una disamina circa la possibilità di ricondurre il c.d. *danno da machine learning* all'interno delle ordinarie categorie di danno da prodotto.

Diversamente, nella categoria dei *danni derivanti da scelte* verranno analizzate situazioni riconducibili al c.d. *Trolley problem*, situazioni dilemmatiche ed eticamente complesse in cui il programmatore dell'algoritmo viene posto ex ante dinanzi a un dilemma etico

di particolare complessità in cui ogni possibile opzione comporti inevitabilmente il generarsi di un danno. Si tratta di situazioni in cui appare ancora più lampante lo stretto legame esistente tra la risoluzione della questione sotto un profilo giuridico e la risoluzione della stessa in senso etico. È per questo motivo che prima di affrontare nello specifico questioni di carattere prettamente penalistico si è ritenuto necessario discutere circa la possibilità di risolvere, in via teorica, tale dilemma adottando un approccio di carattere consequenzialistico o, al contrario, deontologico; passando poi a riflessioni relative all'individuazione del soggetto sul quale dovrebbe ricadere l'ardua scelta. Si è discusso in questo senso circa l'opportunità di delegare la risoluzione del c.d. Trolley problem alternativamente alle società produttrici dei prodotti di A.I., agli stessi conducenti e infine, l'ipotesi che sembrerebbe essere più accreditata, direttamente allo stato tramite la legge, analizzando così per ciascuna di tali alternative i relativi vantaggi e le relative problematiche.

Nel terzo capitolo verrà infine affrontata la tematica più specifica, e sicuramente più complessa, relativa al problema dell'allocazione della responsabilità penale per i danni provocati dai prodotti dell'A.I. L'analisi verrà condotta partendo da un doveroso parallelismo tra le proposte elaborate in sede Europea con riferimento all'allocazione della responsabilità in ambito civile e le moderne dottrine penalistiche sviluppatasi in tempi recentissimi intorno a tali temi, comparandole e affiancandole continuamente agli insegnamenti dei grandi filosofi del passato. Si tratta in realtà di un *modus operandi* non casuale, ma piuttosto doveroso anche al fine di far emergere le cause della maggior complessità delle problematiche attinenti all'allocazione della responsabilità penale in simili situazioni, e dunque, al fine di individuare la strada ideale che il legislatore potrebbe seguire al fine di conciliare istanze, principi, diritti e doveri diversi e spesso in contrasto tra loro. Si analizzerà il problema della responsabilità penale distinguendo, così come si è già fatto nel secondo capitolo, le cause e le fonti da cui potrebbe scaturire il danno: malfunzionamento dell'algoritmo di machine learning e scelte operate in sede di programmazione dall'*uomo dietro la macchina*. Si tratta di una distinzione che risulta essere fondamentale in vista dell'individuazione del soggetto responsabile in quanto, se nel primo dei due casi esaminati la responsabilità potrebbe ricadere alternativamente sul programmatore, sul produttore ovvero, seguendo gli insegnamenti delle proposte più dirimpenti, direttamente sullo stesso robot; nel caso di danni derivanti da scelte solo colui che ha programmato l'algoritmo potrà realisticamente essere gravato di una qualche forma di responsabilità. In quest'ultimo caso dunque, la discussione non verterà più sulla ricerca del soggetto

responsabile per le azioni poste in essere dal robot, ma piuttosto sulla possibilità di scagionare taluno per aver assunto delle scelte in qualche maniera obbligate.

Più nello specifico, nel caso in cui il danno sia provocato da un'evoluzione anomala dell'algoritmo si affronterà innanzitutto il tema relativo all'allocazione della responsabilità penale a seconda del grado di autonomia del prodotto dell'A.I., e quindi dei diversi livelli di automazione delle automobili driverless. Si ipotizzerà nello specifico, per le vetture non ancora completamente autonome, un'estensione della tradizionale responsabilità del conducente umano, il quale assumerà la veste di *conducente potenziale*. Non si potrà tuttavia far a meno di notare come una simile soluzione appaia intessuta di numerosi profili di criticità tra i quali emerge il c.d. *control dilemma* secondo il quale, così facendo, si giungerebbe ad una responsabilizzazione onnicomprensiva del conducente potenziale, che non verrebbe mai liberato da responsabilità anche in caso di eventualità lesive imprevedibili e ingovernabili, conducendo dunque a un contrasto con il principio di colpevolezza, pilastro cardine del diritto penale classico.

Si è poi passati ad una disamina dell'allocazione della responsabilità nel caso in cui i danni siano stati provocati da prodotti completamente automatizzati: errori che, seppur numericamente esigui, non potranno essere imputati a un "conducente", figura assente in un simile contesto. Si analizzeranno pertanto le due opposte strade intraprese dalla dottrina: in primo luogo l'ipotesi, sicuramente più audace e fantasiosa, dell'attribuzione della responsabilità penale direttamente al prodotto dell'intelligenza artificiale, analizzando di conseguenza le ragioni a sostegno di tale teoria e quelle contrarie; in secondo luogo si analizzerà la proposta più accreditata in dottrina di un'estensione della categoria penalistica del danno da prodotto difettoso anche agli eventi lesivi cagionati da tali agenti. Tuttavia, nell'analizzare questa seconda strada individuata dalla dottrina, non si potrà prescindere dal riportarne ed esaminarne i relativi riscontri problematici dovuti al fatto che il *prodotto intelligente* appare indiscutibilmente più complesso rispetto ai tradizionali prodotti oggetto della normativa sulla responsabilità da prodotto difettoso. Le auto a guida autonoma sono infatti dotate della capacità di agire e apprendere nel tempo: caratteristica idonea a rendere infinitamente più complessa la ricostruzione della trama causale del comportamento della macchina, nonché la misura della diligenza richiesta al produttore, apportando di conseguenza un'ineliminabile componente di imprevedibilità nel comportamento degli A.I.

Diverso è il caso in cui il danno sia stato provocato da una scelta assunta ex ante in fase di programmazione in una di quelle situazioni che impongano necessariamente il sacrificio di una o più vite a discapito di altre. Al fine di escludere la responsabilità penale del programmatore si analizzerà dunque la dottrina giuridica della necessità quale promettente base di ragionamento per l'analisi in questione e si sonderà la possibilità di provvedere ad un'estensione dello stato di necessità alla situazione in esame, sia esso considerato alla stregua di una scusante ovvero alla stregua di una causa di giustificazione. Si tratta di una questione di non agevole soluzione stante il fatto che la situazione di pericolo che si presenta in sede di programmazione, nonostante sia inevitabile, appaia priva del requisito dell'attualità richiesto invece dall'articolo 54 c.p., circostanza questa che ne rende difficoltosa l'applicazione.

**Capitolo1:**  
**Contesto di riferimento**

## **1. IMPORTANZA DELL'ANALISI DEL CONTESTO DI RIFERIMENTO.**

### **Dal cibernetista di Platone al giurista della quarta rivoluzione industriale: perchè conoscere il contesto**

*“Un buon governante è come un buon cibernetista (pilota di nave) sa dove vuole andare, conosce la barca, le maree e i venti, giunge a compromessi allungando o modificando la rotta, ma alla fine arriva nel posto giusto”<sup>1</sup>.*

Le parole di Platone appaiono adatte a introdurre il primo capitolo di tale analisi.

Seppur espresse in un'epoca in cui il problema dell'Intelligenza Artificiale e della sua regolamentazione non era nemmeno lontanamente immaginabile, il suo pensiero sembra invece cogliere la reale essenza della questione.

L'obiettivo che tale tesi si propone di raggiungere è quello di analizzare il problema della responsabilità penale per i danni derivanti dai diversi prodotti dell'intelligenza artificiale e di coglierne così le relative implicazioni sul piano etico-giuridico.

Per agire quindi come il buon governante di Platone, che per arrivare alla meta *deve conoscere la barca, le maree e i venti*, risulta essenziale partire da un'analisi dell'Intelligenza Artificiale e delle sue principali applicazioni, tra cui emergono le automobili a guida autonoma. Solo così sarà possibile arrivare nella *giusta direzione*, ossia inquadrare correttamente i problemi della responsabilità penale e le conseguenti ricadute etico-giuridiche.

## **2. INTELLIGENZA ARTIFICIALE.**

### **2.1 Definizione tecnica e considerazioni filosofiche: A.I. matrimonio o divorzio tra intelligenza e capacità di agire?**

L'Intelligenza Artificiale viene definita dal dizionario Di Mauro come *l'insieme di studi e tecniche che tendono alla realizzazione di macchine, specialmente calcolatori elettronici, in grado di risolvere problemi e di riprodurre attività proprie dell'intelligenza umana*<sup>2</sup>.

---

<sup>1</sup> Platone, *Alcibiade primo*, 125 b, 125 c., trad. it. a cura di D. Puliga, Rizzoli, Segrate, 1995. p. 91.

<sup>2</sup> T. De Mauro, *Grande dizionario italiano dell'uso*, voce “intelligenza artificiale”, Torino 2000

Non esiste ancora, tuttavia, una definizione universalmente accettata in quanto l'AI (dall'inglese Artificial Intelligence) è un settore estremamente recente e in fortissima evoluzione.

Tra le definizioni più accreditate si può ricordare quella offerta da Knight per il quale si tratterebbe dello “*studio delle facoltà mentali mediante l'uso di modelli computazionali*”<sup>3</sup> e quella fornita dalla Commissione Europea nel 2018 che definisce Intelligenza Artificiale (IA) come l'insieme dei “*sistemi che mostrano un comportamento intelligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi*”<sup>4</sup>.

In realtà, il primo problema dell'A.I., al quale si riconnettono le difficoltà di offrirne una definizione univoca, risiede proprio nella denominazione stessa. La nozione di Intelligenza artificiale, come sottolineato dal Professor Floridi, rappresenta un ossimoro: *se una cosa è artificiale non è intelligente, di riflesso se una cosa è intelligente non è artificiale*<sup>5</sup>. Con tale locuzione viene attribuito al requisito dell'artificialità qualcosa di essenzialmente naturale: l'intelligenza, la prerogativa più identitaria della natura umana. L'ossimoro è sicuramente provocatorio e idoneo a far sorgere numerosi interrogativi, il primo tra i quali riguarda la possibilità per una macchina di poter effettivamente pensare e ragionare al pari di un essere umano.

Invero il nome di questa innovativa branca dell'informatica fu coniato per la prima volta nel 1956 durante la “Dartmouth Summer Research Project on Artificial Intelligence”, conferenza in cui per la prima volta si presentò al pubblico un oggetto allora del tutto sconosciuto: il computer. Bisognava pertanto servirsi di un termine forte per presentare un prodotto altrettanto rivoluzionario e, difatti, l'idea fu seguita da un enorme successo. Sennonché, come sottolineato poc'anzi, il termine stesso è fuorviante. Esso induce erroneamente a pensare, per citare le parole del Professor Floridi, “*a un matrimonio, un'unione tra la capacità biologica di intendere e la capacità di agire; ossia tra l'intelligenza umana e il prodotto tecnologico*”<sup>6</sup>. Tale affermazione riposa sull'idea secondo la

---

<sup>3</sup> E. Rich, K. Knight, *Artificial Intelligence*, McGraw-Hill, New-York, 1991

<sup>4</sup> Commissione Europea, *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni; L'intelligenza artificiale per l'Europa*, 237 final, Bruxelles, 25/4/2018.

<sup>5</sup> L. Floridi, *Onlife, La vita al tempo del digitale*, Politecnico di Milano, 5 ottobre 2019.

<sup>6</sup> L. Floridi. Incontro con Luciano Floridi, inserito nel progetto "Dalla maschera al robot" con il contributo di Fondazione Cariplo, 9 gennaio 2020.

quale, se una macchina è in grado di raggiungere lo stesso risultato di un essere umano, il processo logico che l'ha guidata sarà il medesimo: di conseguenza anche la fonte, ossia l'intelligenza, sarà la stessa.

In realtà così non è, e le stesse modalità di funzionamento dei prodotti dell'A.I. lo dimostrano. *“Non bisogna confondere l'emulazione con il funzionalismo, in forza del quale la stessa funzione è svolta da differenti sistemi fisici. L'emulazione riguarda i risultati: agenti che si emulano l'un l'altro possono conseguire lo stesso risultato, ma ciò in forza di strategie e metodi completamente diversi”*<sup>7</sup>.

L'intelligenza artificiale si presenta quindi come un *divorzio*, una separazione tra la capacità di agire con successo e l'intelligenza per farlo<sup>8</sup>. Ed è proprio tale separazione ad aver determinato il crollo di tutte le certezze preesistenti: l'uomo ha creato macchine, non intelligenti, ma *in grado di “riprodurre attività proprie dell'intelligenza umana”*, per tornare alla definizione offerta dal dizionario De Mauro.

D'altronde l'anima dell'*automa*, anche nella radice etimologica del nome, è la capacità di agire da sé, di muoversi in modo autonomo e coordinato<sup>9</sup>, ma non anche l'intelligenza necessaria per conseguire tale risultato.

Per comprendere quindi il motivo per cui una macchina non intelligente sia in grado di riprodurre attività proprie dell'intelligenza umana e, di riflesso, per comprendere le modalità stesse di funzionamento dei prodotti dell'A.I., sarà necessario partire dal presupposto secondo cui è il mondo stesso ad essere stato modificato, trasformandosi gradualmente in uno spazio amichevole ai robot<sup>10</sup>.

I prodotti dell'intelligenza artificiale non stanno quindi diventando sempre più intelligenti, ma è il mondo, al contrario, che sta divenendo *un'infosfera*<sup>11</sup> sempre più adatta e accogliente alle limitate capacità di tali macchine. *“Se immergiamo gli agenti artificiali nella loro zuppa digitale li ritroveremo felicemente ronzanti”*<sup>12</sup>.

---

<sup>7</sup> L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Raffaello Cortina Editore, Milano, 2017, p. 161.

<sup>8</sup> *Ivi*, pp. 165 e ss.

<sup>9</sup> A. Carcaterra, *Machinae autonomae e decisione robotica*, in *Decisione Robotica*, a cura di A. Carleo, Il Mulino, Bologna, 2019, p.33

<sup>10</sup> L. Floridi, *La quarta rivoluzione*, cit., p.165 e ss.

<sup>11</sup> Il termine viene utilizzato da Luciano Floridi per indicare la graduale informatizzazione dell'intero sistema sociale, una società sempre più basata e condizionata dalle informazioni a disposizione. Per un approfondimento sul tema si veda L. Floridi, *La quarta rivoluzione*, cit., pp. 27 e ss.

<sup>12</sup> L. Floridi, *La quarta rivoluzione*, cit., p. 155.

Si pensi, partendo da una situazione del tutto comune nella realtà giornaliera, agli scanner presenti in tutte le casse dei supermercati: questi prodotti dell'intelligenza artificiale, nel momento in cui vengono passati sopra i prodotti da noi acquistati, sono in grado di rilevare il tipo di prodotto e il relativo prezzo. Il risultato da questi raggiunto tuttavia non è dovuto al fatto che gli scanner siano stati dotati della medesima capacità di discernimento degli esseri umani, ma piuttosto in quanto è la confezione del prodotto stesso ad essere stata modificata, è stata infatti dotata di un codice a barre, un linguaggio comprensibile al robot.

Allo stesso modo, le automobili self-driving, sulle quali si ritornerà in seguito<sup>13</sup>, saranno in grado di accompagnarci in ufficio, ma non in quanto dotate di un'abilità di guida paragonabile a quella umana, piuttosto in quanto l'ambiente stesso verrà modificato, consentendo così alla macchina i suoi spostamenti. Il concetto di automobile così come ora concepito verrà stravolto: il volante o il freno non saranno più necessari per poter realizzare una manovra ma, a risultare indispensabili, saranno piuttosto i sistemi di radar e lidar.

È in quest'ottica che vanno analizzate le relative questioni etiche e giuridiche che seguiranno nei capitoli successivi.

Bisogna rifuggire pertanto da realtà apocalittiche prospettate dai film e dai libri di fantascienza, realtà nelle quali i robot, macchine pensanti e autonome rispetto all'uomo, riescano a prendere, in un futuro più o meno recente, il sopravvento sulla stessa razza umana che li ha ideati e creati.

Piuttosto, nel momento in cui si esaminerà la tematica della responsabilità penale derivante dalle azioni poste in essere dal robot e le questioni etiche connesse all'attribuzione della stessa in un senso piuttosto che un altro, si dovrà sempre tener presente che i prodotti dell'intelligenza artificiale non sono dotati un'intelligenza biologica paragonabile a quella umana ma, più semplicemente, riescono a raggiungere un determinato risultato per il solo fatto che noi stessi abbiamo progressivamente modificato l'intero ambiente al fine di permettere a tali macchine di operare, e di operare a servizio dell'uomo<sup>14</sup>.

---

<sup>13</sup> Si veda il cap. 1, par. 3, pp. 33 e ss. di questo lavoro.

<sup>14</sup> L. Floridi, *La quarta rivoluzione*, cit., pp. 165 e ss.

## 2.2 Modalità di funzionamento

È doveroso sottolineare che non si vuole affrontare in tale sede una dettagliata ricognizione delle modalità di funzionamento dei sistemi di intelligenza artificiale dal punto di vista tecnico, in quanto, seppur strettamente collegata con l'ambito delle tecnologie e dell'informatica, l'analisi in questione è mirata ad affrontare problemi di carattere prettamente etico-giuridico.

Tuttavia, per riprendere il concetto espresso dalle parole di Platone, al fine di comprendere a fondo come la materia possa in futuro essere regolata, in particolar modo dal punto di vista penalistico, appare necessario rilevare, almeno in via generale, i principi base di funzionamento dell'A.I. in modo da percepire la reale portata e i possibili approdi di tale scienza.

Per comprendere, se non altro in un'ottica generale, le modalità di funzionamento dell'A.I. risulta necessario focalizzare l'attenzione su due concetti ricorrenti: Intelligenza artificiale e *Machine learning*.

Non sempre chiara tuttavia è la relazione intercorrente tra tali tecnologie. A tal proposito si fa presente che in molti contesti divulgativi tali termini sono spesso erroneamente utilizzati alla stregua di concetti interscambiabili tra loro. In realtà il rapporto tra gli stessi è più complesso. Usando la simbologia degli insiemi, Artificial Intelligence (AI), Machine Learning (ML) possono essere considerati come due ambiti scientifici racchiusi l'uno nell'altro<sup>15</sup>.

### INTELLIGENZA ARTIFICIALE

L'Intelligenza Artificiale, termine più ampio per definire qualsiasi programma che agisca secondo una logica razionante, è la scienza che sviluppa l'architettura necessaria affinché le macchine funzionino come il cervello umano, consentendo così di imparare dall'esperienza, di adeguarsi alle nuove informazioni ricevute e dunque svolgere compiti simili a quelli dell'uomo. È, detto in altri termini, l'obiettivo da raggiungere.

---

<sup>15</sup> A tali tecnologie se ne aggiunge in realtà una terza rappresentata dal Deep Learning. Questa è la scienza concettualmente più ristretta e rappresenta una particolare modalità di machine learning. Per un approfondimento sul tema si veda F. Collet, *Deep learning con Python. Imparare a implementare algoritmi di apprendimento profondo*, Apogeo, Milano, 2020.

## MACHINE LEARNING

Il Machine Learning, o apprendimento automatico, rientra nel più ampio concetto di Intelligenza Artificiale. È rappresentato dall'insieme di tutte quelle tecniche e metodi tramite cui si permette alla macchina, utilizzando l'esperienza, di svolgere compiti che non siano stati già precedentemente programmati e migliorando così, col tempo, le proprie capacità, le proprie risposte e funzioni<sup>16</sup>. Una macchina sarà pertanto in grado di imparare a svolgere una determinata azione anche se la stessa non è mai stata programmata tra le azioni possibili.

Ciò risulta possibile in quanto alla base dell'apprendimento automatico sono presenti una serie di differenti algoritmi<sup>17</sup> che, partendo da nozioni primitive rappresentate da dati, sapranno prendere una specifica decisione piuttosto che un'altra, agire in un modo piuttosto che in un altro.

Per comprendere a pieno il funzionamento e le relative ricadute etico-giuridiche appare necessario dunque operare una distinzione tra gli algoritmi tradizionali, o basati su regole, e gli algoritmi di machine learning.

Nei primi, le istruzioni sono definite da un essere umano. Si tratta di istruzioni dirette e prive di ambiguità: fare A; se accade A allora fare B e così via. Si tratta di algoritmi che rendono prevedibile ogni azione e decisione della macchina. Di conseguenza, ogni azione potenzialmente dannosa risulterà imputabile a colui che abbia progettato l'algoritmo stesso<sup>18</sup>.

Questa affermazione risulta valida anche nei casi di algoritmi più complessi e indecifrabili frutto di anni di lavoro. Nel 2013 infatti a seguito di un incidente mortale, causato da un'accelerazione improvvisa ed incontrollabile di un veicolo Toyota, la casa automobilistica fu condannata a pagare un risarcimento di tre milioni di dollari. La responsabilità della stessa fu dichiarata in quanto si attribuì la causa della tragedia ad un'istruzione errata sepolta sotto un ammasso caotico di software<sup>19</sup>.

---

<sup>16</sup> A. De Mauro, *Big data analytics: guida per iniziare a classificare e interpretare dati con il machine learning*, Apogeo, Milano, 2019, p. 69.

<sup>17</sup> Per algoritmi si intende una procedura composta da una serie di istruzioni per la risoluzione di un problema o l'esecuzione di un compito, soprattutto da parte di un computer. Definizione fornita da J. Brill, *Data Brokers: A Call for Transparency and Accountability*, in Reports of Federal Trade Commission, maggio 2014, pp. 11.

<sup>18</sup> H. Fry, *Hello World. Essere umani al tempo del digitale*, Bollani Boringhieri, Torino, 2019, p. 19.

<sup>19</sup> Per un approfondimento sul tema si veda P. Koopman, *A case study of Toyota unintended acceleration and softer safety*, Carnegie Mellon University, Pittsburg, 2014.

Gli algoritmi di Machine learning sono invece guidati da una logica diversa e più complessa in cui non sempre risulta possibile attribuire la causa di eventuali malfunzionamenti a colui che abbia progettato l'algoritmo. Questi infatti si basano sullo sviluppo delle reti neurali artificiali<sup>20</sup>, ispirandosi di fatto all'apprendimento degli esseri viventi.

Si forniscono alla macchina dei dati, un obiettivo e un feedback, rinforzando così i comportamenti corretti e scoraggiando quelli sbagliati<sup>21</sup>. Con l'esperienza accumulata la macchina sarà così in grado di comprendere il modo migliore per raggiungere l'obiettivo, agendo efficacemente e *“riproducendo le attività proprie dell'intelligenza umana”*<sup>22</sup>.

L'utilizzo di tali algoritmi è sempre più frequente soprattutto in quegli ambiti in cui non sarà sufficiente stilare una lista di istruzioni per far eseguire alla macchina una determinata azione e risolvere di conseguenza i problemi che si prospettano. Non sarà ad esempio sufficiente una lista di istruzioni per eseguire correttamente una traduzione, riconoscere un oggetto o una voce, guidare autonomamente per chilometri o centrare un obiettivo militare riconoscendo, e di conseguenza risparmiando, i civili.

### **2.3 Self learning e introduzione delle questioni giuridiche e morali**

Si analizzeranno in seguito le problematiche derivanti da una simile modalità di funzionamento. In questo paragrafo si ritiene sufficiente far presente che, nel momento in cui si consente alla macchina di trovare da sé la soluzione per raggiungere un determinato risultato, in molti casi questa sceglierà una strada che, anche agli occhi dei programmatori più brillanti, risulterà incomprensibile<sup>23</sup>.

Infatti, anche se una macchina può ottenere risultati analoghi a quelli di un essere umano, il suo modo di operare e la sua logica sono completamente diverse da quelle umane: può

---

<sup>20</sup>L'utilizzo delle c.d. Reti Neurali Artificiali è ciò che ha permesso ai sistemi intelligenti di migliorare sempre più le diverse capacità di apprendimento. Esse rappresentano un algoritmo, già ideato verso la fine degli anni 60, in grado di riprodurre ragionamenti tipici degli esseri umani nelle differenti situazioni. Si tratta in particolare di modelli matematici composti da “neuroni” sintetici, che emulano le reti neurali biologiche e che vengono utilizzate per risolvere problemi attraverso l'A.I. Pertanto, così come la rete neurale biologica anche la rete neurale artificiale percepisce degli stimoli (in quest'ultimo caso ciò è reso possibile da tecniche come radar, lidar, gliss, infrared, visione artificiale e così via) e agisce di conseguenza anche sulla base dell'esperienza, ossia delle informazioni già presenti nel sistema di partenza o acquisite tramite precedenti utilizza. Per una disamina più dettagliata si veda P. Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, New York, 2013, pp. 93 e ss.

<sup>21</sup> H. Fry, *Hello World. Essere umani al tempo del digitale*, cit. p. 20.

<sup>22</sup> Per tornare alla definizione fornita dal Dizionario De Mauro.

<sup>23</sup> H. Fry, *Hello World. Essere umani al tempo del digitale*, cit., p. 20.

riconoscere un'immagine, un pedone, un civile o fare un errore, ma questo avviene tramite processi completamente diversi da quelli che inducono in inganno un essere umano.<sup>24</sup> Ad esempio, un gruppo di ricercatori giapponesi ha dimostrato come fosse sufficiente oscurare un singolo pixel di un'immagine raffigurante un neonato per far sì che l'algoritmo di machine learning smettesse di classificarla correttamente e cominciasse piuttosto a riconoscervi un gatto<sup>25</sup>.

L'esempio sopra citato apre gli occhi su una problematica molto più ampia. Nel caso in cui lo stesso errore venga commesso da una macchina in grado di produrre ingenti danni, come un'automobile driverless o un drone militare, bisognerà non solamente individuare il soggetto responsabile ma, prima ancora, individuare se sia possibile punire o meno qualcuno per il fatto commesso dalla macchina. Si è in presenza di ciò che il Parlamento Europeo indica come *responsability gap*<sup>26</sup>.

Come si approfondirà più avanti nel corso della trattazione, nel capitolo relativo all'ethic by design, la parte più delicata del suddetto processo è proprio la creazione del modello di apprendimento della macchina, il quale viene costruito e messo in moto a partire dai dati<sup>27</sup>. Gli algoritmi di machine learning infatti estrapolano, analizzano, elaborano e mettono in relazione un'enorme massa di dati<sup>28</sup> al fine di emettere la c.d. *decisione robotica*, la risposta più adeguata a far fronte a una specifica situazione.

La decisione robotica si potrebbe infatti definire come il connubio tra tre diversi elementi: i dati forniti dai sensori, i modelli della realtà<sup>29</sup> e gli algoritmi di decisione<sup>30</sup>. Si tratta di tre elementi che, al pari di quanto avviene nel processo decisionale degli esseri umani, si prestano ad essere influenzati da una connotazione soggettiva portando di conseguenza a distorsioni e, cosa più significativa, ad errori.

---

<sup>24</sup>G. F. Italiano, *Le sfide interdisciplinari dell'intelligenza artificiale*, in «Analisi Giuridica dell'Economia, Studi e discussioni sul diritto dell'impresa», 1, 2019, p. 13.

<sup>25</sup>J. Su, D. Vasconcellos Varagas e K. Sakurai, *One Pixel attack for fooling deep neural networks*, in «IEEE Transactions on Evolutionary Computation», 5, 2018, pp. 828 e ss.

<sup>26</sup>Parlamento Europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, Risoluzione del 16 febbraio 2017, par. introduttivo.

<sup>27</sup>Così G. F. Italiano, *Le sfide interdisciplinari dell'intelligenza artificiale*, cit., p. 13.

<sup>28</sup>A tal proposito si fa spesso riferimento ai Big Data. Il termine Big Data viene utilizzato per descrivere sia ampi e complessi set di dati sia l'insieme di tutte le tecnologie e le metodologie di analisi di dati massivi in grado di estrapolare, analizzare e mettere in relazione un'enorme mole di dati eterogenei, strutturati e non strutturati, per scoprire i legami tra fenomeni diversi e prevedere quelli futuri. Per un approfondimento sul tema si veda A. De Mauro, *Big data analytics*, cit., pp. 1-11.

<sup>29</sup>Ove per modelli della realtà si fa riferimento a un insieme di relazioni che rappresentano dei vincoli o delle regole comportamentali cui gli A.I. devono necessariamente sottostare. Carcaterra porta a tal proposito l'esempio delle self driving car, le quali devono ubbidire non solo alle leggi della fisica ma anche a quelle del codice della strada. A. Carcaterra, *Machinae autonomae e decisione robotica*, cit., pp. 43 e ss.

<sup>30</sup>A. Carcaterra, *Machinae autonomae e decisione robotica*, cit., p. 38.

Nel paragrafo precedente si è discusso di come gli algoritmi di machine learning permettano alla macchina di apprendere e svolgere funzioni per le quali non sia stata ab origine programmata grazie all'esperienza, vale a dire alla mole di dati che la macchina dovrà analizzare ed elaborare. I dati tuttavia sono solitamente soggetti a diversi difetti tra i quali si cita l'incompletezza, la distorsione e l'incertezza. Da qui la domanda: quand'è che un dato ottenuto costituisce effettivamente una buona rappresentazione della realtà?<sup>31</sup>

In generale, nel processo decisionale biologico o comunque tipico degli esseri umani, la descrizione della realtà di fatto, sebbene spesso percepita come oggettiva, se ben analizzata, può infatti presentare segni di soggettività nel momento in cui un determinato stato di fatto viene trasformato in un'informazione elaborabile, ossia portatrice di un contenuto informativo definito. Pertanto, se le decisioni cui giungiamo sono conseguenza dell'osservazione così come ci viene rappresentata attraverso questi dati, dovremmo desumere che le decisioni stesse porteranno i segni dell'arbitrarietà e della soggettività che risiede all'origine della descrizione dei fatti<sup>32</sup>.

Si tratta di un ragionamento che resta valido anche per la c.d. *decisione robotica*: La tipologia e la qualità dei dati forniti, nonché la quantità degli stessi, comporteranno pertanto significative conseguenze sul funzionamento e sulle scelte operative di tali prodotti. Da un lato, se i dati forniti originariamente, o quelli acquisiti con l'esperienza, non sono sufficientemente rappresentativi si potrebbe sviluppare il rischio che le decisioni prese dalla macchina siano potenzialmente dannose e pregiudizievoli. Le conseguenze che ne derivano appaiono destinate a riflettersi sul piano della responsabilità in quanto, pur dipendendo da una scelta autonoma della macchina, risultano indirettamente prevedibili e riconducibili a soggetti determinati. Del tema si tratterà nello specifico nel capitolo 2 al paragrafo sull'*ethic by design*.

Dall'altro lato, l'Intelligenza Artificiale si sviluppa grazie alla quantità dei dati a disposizione: più la macchina ne conosce, più ne elabora e meglio sarà in grado di moltiplicare le sue funzioni e le sue potenzialità. Le conseguenze, in questo caso, avranno un impatto non sul piano della responsabilità ma piuttosto sul piano economico. La tematica non sarà oggetto della trattazione in esame, ma si ritiene per completezza espositiva, in quanto legata al futuro della regolamentazione dei dati personali, farne un breve cenno.

---

<sup>31</sup> *Ivi*, p. 42.

<sup>32</sup> *Ivi*, p. 43.

È stata ribadita da un recente studio del magazine «Forbes» la superiorità della Cina nel campo dell'intelligenza artificiale, paese in cui ben l'85% delle aziende è impegnata a sviluppare forme più o meno avanzate di intelligenza artificiale. Nella Cina Popolare la rete, e l'acquisizione di dati, è controllata da un unico attore: lo Stato. Si tratta pertanto di uno strumento politico-economico nelle mani esclusive del governo e di una legislazione che non tiene doverosamente in considerazione la privacy nonché i diritti digitali degli individui. In un paese in cui è possibile disporre senza restrizioni, e in tempo reale, di tutti i dati prodotti da oltre un miliardo di abitanti, lo sviluppo l'intelligenza artificiale ad una velocità senza rivali costituisce la logica conseguenza<sup>33</sup>.

L'Unione Europea, al contrario, si sta orientando verso un approccio diametralmente opposto, basato sulla protezione dei diritti digitali, dalla riservatezza alla proprietà intellettuale<sup>34</sup>. La strada seguita dal legislatore di Strasburgo mira pertanto a contemperare le esigenze legate alla privacy con il perseguimento dello sviluppo tecnologico. Ma questo è destinato a riflettersi inevitabilmente sulla competitività nel settore dell'intelligenza artificiale. Il rischio è che i paesi dell'Unione Europea potrebbero trovarsi presto a divenire consumatori di soluzioni sviluppate altrove che non rispondono ai valori e ai principi etici dettati dalla Comunità<sup>35</sup>. Da qui la necessità di trasferire lo Stato di Diritto nel cyberspazio per consentirgli di governare e di coesistere con l'intelligenza artificiale.

Infine, e in aggiunta alla problematica relativa all'incompletezza dei dati forniti alla macchina, vi è un altro fattore tale da condurre il prodotto dell'A.I. a prendere decisioni indesiderate e pregiudizievoli. Si fa riferimento in particolare al c.d. *curse of dimensionality*: l'esplosione di scenari futuri, ipotetici e potenzialmente incontrollabili che l'algoritmo di machine learning dovrà necessariamente tenere in considerazione al fine di adottare la decisione ottimale<sup>36</sup>. L'agente razionale infatti, in base all'esperienza passata e nel rispetto di alcuni vincoli attuali, agisce in modo tale da massimizzare l'utilità attesa, ossia in modo tale da prendere una determinata decisione guardando all'effetto o alle conseguenze che tale decisione produrrà<sup>37</sup>. Tuttavia, le conseguenze di un'azione sono

---

<sup>33</sup> M. Molinari, *Assedio all'Occidente*, La Nave di Teseo Editore, Milano, 2019, pp. 22 e ss.

<sup>34</sup> Si tratta dell'emanazione del *GDPR (General Data Protection Regulation)* del Parlamento Europeo e del Consiglio del 27 aprile 2016, approvato con Regolamento UE 2016/679 e applicabile a decorrere dal 25 maggio 2018.

<sup>35</sup> Le preoccupazioni sono state manifestate in particolare dalla Commissione Europea nel documento *Intelligenza artificiale per l'europa; Comunicazione della commissione al parlamento europeo, al consiglio, al comitato economico e sociale europeo e al comitato delle regioni*, Bruxelles, 25.4.2018.

<sup>36</sup> R. E. Bellman, *Applied Dynamic Programming*, Princeton-Oxford, 2010. Cfr A. Carcaterra, *Machinae autonomae e decisione robotica*, cit., p. 47.

<sup>37</sup> A. Carcaterra, *Machinae autonomae e decisione robotica*, cit., p. 45.

temporalmente collocate in un momento successivo a quello in cui la decisione viene assunta, e ciò implica un necessario coordinamento tra l'azione del robot e la valutazione di eventi futuri tramite l'adozione di un modello della realtà di tipo predittivo, anche rispetto a orizzonti temporali significativamente estesi. È in tale circostanza che risiede l'ulteriore aspetto problematico della decisione robotica: nelle programmazioni a lungo termine, i possibili scenari da tenere in considerazione sono destinati ad aumentare in modo incontrollabile dando così luogo a quel *curse of dimensionality* di cui si è detto sopra<sup>38</sup>.

## 2.4 La quarta rivoluzione industriale

L'analisi delle modalità di funzionamento dell'intelligenza artificiale ha messo in luce un ulteriore aspetto di rilevanza non secondaria: la società sta facendo il suo ingresso nell'era della *quarta rivoluzione*.

Di ciò ne sembra dar conferma anche la stessa Commissione Europea che, nella recente risoluzione fornita al Parlamento Europeo nel maggio 2018, scrive: “*Siamo agli albori di una quarta rivoluzione industriale: l'era, preannunciata dal genio di Alan Turing nella quale esseri umani e macchine intelligenti interagiranno sempre più strettamente. Macchine sempre più interconnesse, autonome e capaci di pensare*”<sup>39</sup>

L'espressione “quarta rivoluzione”, coniata da Klaus Schwab, economista tedesco, indica una “*sempre crescente compenetrazione tra mondo fisico, digitale e biologico in grado di generare uno stravolgimento radicale del modo in cui viviamo, lavoriamo, ci relazioniamo*”<sup>40</sup>. In altre parole, e cercando di riassumere in poche righe il pensiero espresso da Schwab, questa rivoluzione viene innescata dall'utilizzo, all'interno delle imprese, delle c.d. *tecnologie abilitanti* (quali robotica avanzata, A.I, Internet of things, stampa 3D e l'analisi dei Big Data)<sup>41</sup>. In quanto risultato di questa tempesta perfetta di tecnologie, la

---

<sup>38</sup> *Ivi*, p. 47.

<sup>39</sup> Commissione Europea, *Relazione della Commissione al Parlamento Europeo, al Consiglio e al Comitato Economico e Sociale Europeo sull'applicazione della direttiva del Consiglio relative al ravvicinamento delle disposizioni legislative, regolamentari ed amministrative degli Stati Membri in materia di responsabilità per danno da prodotti difettosi* (direttiva 85/374/CEE), 7 maggio 2018.

<sup>40</sup> K. Schwab, *La quarta rivoluzione industriale*, Franco Angeli, Milano, 2016, p. 12.

<sup>41</sup> B. Marr, *The 4th Industrial Revolution Is Here - Are You Ready?*, in «Forbes», Agosto 2018.

quarta rivoluzione industriale sta spianando la strada ai cambiamenti che avverranno nel modo in cui viviamo e sta sconvolgendo radicalmente quasi tutti i settori economici. Tutto questo sta accadendo a un ritmo vorticoso mai visto prima<sup>42</sup>.

Come scrive lo stesso Schwab, la quarta rivoluzione ha, tanto il potenziale per migliorare la qualità della vita per la popolazione di tutto il mondo, quanto quello di comprometterla. *“I cambiamenti sono così tanto profondi che nella storia dell’uomo non si è mai assistito a un periodo così promettente ma altrettanto potenzialmente pericoloso, considerando la sua idoneità a stravolgere i mercati del lavoro”*<sup>43</sup>.

Da un lato pertanto, la quarta rivoluzione condurrà inevitabilmente ad una disuguaglianza che potremmo definire *ontologica*, separando chi si adatta e chi si oppone. I cambiamenti significativi su aspettative di vita, salute, conoscenze e abilità nell’uso delle tecnologie, porteranno e ridefinire lo stesso concetto di essere umano. La sfida sembra dunque risiedere nella capacità di controllare simili processi, ad esempio favorendo la capacità di empatia e di acquisire competenze sociali<sup>44</sup>.

Dall’altro lato, non può essere sottovalutato il fatto per cui, come le precedenti rivoluzioni industriali, la quarta rivoluzione ha in sé il potenziale per stravolgere gli equilibri economici mondiali.

Pertanto, così come a seguito dell’avvento delle macchine a vapore le imprese che ne erano prive persero la loro competitività, ad oggi, i paesi e le aziende che non percepiscano, e di conseguenza non sfruttino, la reale portata innovativa delle nuove *tecnologie abilitanti*, rischiano di perdere il proprio posto all’interno di un mercato sempre più globalizzato.

È per questa ragione che nel 2016 è stato emanato in Italia dal Ministero dello sviluppo economico il *Piano Nazionale Industria 4.0*, nella cui introduzione, redatta dall’allora ministro dello sviluppo economico Calenda, si legge che *“Il piano investe tutti gli aspetti del ciclo di vita delle imprese che vogliono acquisire competitività, offrendo un supporto negli investimenti, nella digitalizzazione dei processi produttivi, nella valorizzazione*

---

<sup>42</sup> K. Schwab, *La quarta rivoluzione industriale*, cit., pp. 30-37.

<sup>43</sup> Ivi, p. 37.

<sup>44</sup> Il tema viene approfondito da S. Turkel, *La conversazione necessaria. La forza del dialogo nell’era digitale*, Einaudi, Torino, 2016.

*della produttività dei lavoratori, nella formazione di competenze adeguate e nello sviluppo di nuovi prodotti e processi*”<sup>45</sup>.

Ma l’obiettivo di mantenere la competitività nel nuovo mercato globale è avvertito anche in sede Europea, la cui linea d’azione emerge chiara in numerosi documenti.

Tra questi si menziona la Comunicazione della Commissione “*L’intelligenza artificiale per l’Europa*” adottata nell’Aprile 2018, nella quale, dopo aver sottolineato come l’A.I. sia ormai una realtà, rileva come, dinnanzi a una accanita concorrenza globale, sia necessario sviluppare un solido quadro di riferimento europeo<sup>46</sup>.

L’UE, fa notare la commissione, “*ha in sé il potenziale per assumere un ruolo guida nello sviluppo e nello sviluppo e nell’impiego dell’intelligenza artificiale sulla base dei valori condivisi dall’intera Comunità*”<sup>47</sup>.

Se così non fosse, ossia se l’Unione Europea non si attivasse in tal senso, “*rischierebbe di perdere le opportunità offerte dall’A.I. con la prospettiva di una fuga di cervelli e di divenire un consumatore di soluzioni sviluppate altrove*”<sup>48</sup>, soluzioni che potrebbero rispecchiare principi diversi da quelli condivisi a livello Europeo.

Il medesimo punto di vista viene altresì condiviso dal Parlamento Europeo nella più recente risoluzione del Febbraio 2019 “*Una politica industriale europea globale in materia di robotica e intelligenza artificiale*” nella quale vengono affrontati diversi temi correlati alle nuove tecnologie<sup>49</sup>. Tra questi risalta in particolar modo quello relativo al mutamento della società a seguito dell’avvento delle nuove tecnologie e della robotica.

Il Parlamento Europeo, dopo aver messo in luce la circostanza secondo la quale, l’automazione, unita all’intelligenza artificiale, incrementerà la produzione e la produttività, osserva tuttavia come tale mutamento porterà a uno stravolgimento delle ordinarie qualifiche lavorative<sup>50</sup>. Esorta pertanto gli Stati membri non solo a concentrarsi sulla riqualificazione dei lavoratori dei settori industriali più colpiti dall’avvento dell’automazione

---

<sup>45</sup> Ministero dello sviluppo economico, *Piano Nazionale Industria 4.0*, 2016, introduzione.

<sup>46</sup> Commissione Europea, *Intelligenza artificiale per l’Europa*, cit., pp. 1-2.

<sup>47</sup> Sempre Commissione Europea, *Intelligenza Artificiale per l’Europa*, cit, p. 2. Per “Valori condivisi dalla comunità” si fa riferimento in particolare allo “*sviluppo di un quadro adeguato che promuova l’innovazione e rispetti i valori fondamentali dell’Unione e i diritti fondamentali, oltre che ai principi etici come la responsabilità e la trasparenza*”

<sup>48</sup> Commissione Europea, *Intelligenza artificiale per l’Europa*, cit., p. 7.

<sup>49</sup> Parlamento Europeo, *Una politica industriale europea globale in materia di robotica e intelligenza artificiale*, 12 febbraio 2019, 2018/2088(INI).

<sup>50</sup> *Ivi*, p. 4.

delle mansioni, ma anche a considerare adeguatamente gli aspetti riguardanti l'istruzione, sottolineando a tal proposito la necessità di includere in tutti i programmi di insegnamento, fin dai primi anni scolastici, le competenze digitali<sup>51</sup>.

Lo scopo perseguito dall'UE è quindi quello di formare una nuova classe di lavoratori pronti a confrontarsi con le tecnologie abilitanti e in grado di poter svolgere le nuove mansioni lavorative sviluppatesi con l'avvento dell'A.I.

L'ordinamento eurounitario mira tuttavia non solo ad adattarsi al cambiamento apportato con la quarta rivoluzione industriale, ma altresì ad assumere un ruolo di guida all'interno del settore dell'intelligenza artificiale. È opportuno far luce pertanto sulla circostanza per cui i diversi documenti emanati non si focalizzano esclusivamente su strategie di carattere tecnico, economico ed organizzativo, ma sottolineano, anche e soprattutto, l'esigenza di lavorare su un concetto di innovazione che possa essere considerato etico ed affidabile.

Questo è quanto esposto nelle “*Ethic Guidelines for Trustworthy A.I.*” redatte da un apposito Gruppo di Alto livello di Esperti nel dicembre 2018 e adottate dal mese di marzo 2019<sup>52</sup>.

Le Linee guida partono dall'assunto secondo cui un approccio etico e affidabile costituisce la via principale per migliorare la competitività dell'UE nel settore dell'intelligenza artificiale. L'idea di fondo è rappresentata dalla circostanza secondo la quale, aumentando la fiducia negli utenti nei confronti di tali prodotti, vi seguirà una maggior diffusione dell'A.I.

Si legge infatti come “*L'affidabilità costituisca un prerequisito per chiunque voglia sviluppare, implementare e utilizzare i sistemi di intelligenza artificiale. Senza sistemi di A.I. manifestamente degni di fiducia, deriverebbero conseguenze indesiderate e la loro diffusione potrebbe essere ostacolata, impedendo così la realizzazione del potenziale sociale ed economico offerto da tali prodotti*”<sup>53</sup>.

---

<sup>51</sup> *Ibidem*.

<sup>52</sup> The High-Level Expert Group in A.I., *Ethic Guidelines for Trustworthy A.I.*, in «Reports and studies of European Commission», Aprile 2019. Nella stessa direzione si muove la Commissione Europea nel *LIBRO BIANCO Sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia*, Bruxelles, 65 final, 19.2.2020, introduzione.

<sup>53</sup> The High-Level Expert Group in A.I., *Ethic Guidelines for Trustworthy A.I.*, cit., p. 4.

Al fine poi di garantire che i sistemi di A.I. possano essere considerati affidabili il documento fissa sette valori generalmente condivisi<sup>54</sup> e individua specifici metodi, tecnici e organizzativi, aventi lo scopo di implementarli per tutto il ciclo di vita dei prodotti dell'intelligenza artificiale.

## 2.5 Le diverse applicazioni dell'A.I.

La quarta rivoluzione industriale è pertanto alle porte e di ciò se ne trova conferma non solamente nelle pagine dei documenti e progetti istituzionali ma, più semplicemente, osservando con maggiore attenzione la realtà in cui viviamo.

Dopo il primo vero successo dell'intelligenza artificiale, rappresentato dalla vittoria avvenuta nel 1996 della macchina *Deep Blue*, realizzata dall'IBM, sul campione di scacchi allora in carica Garry Kasparov<sup>55</sup>, si può pacificamente constatare che attualmente l'A.I. e il suo utilizzo siano oggi molto più reali e radicati di quanto ci si possa immaginare. Vengono infatti ampiamente utilizzati in diversi settori della vita quotidiana, tanto che spesso la percezione è quella di trovarsi di fronte a un comune strumento elettronico piuttosto che ad un prodotto dell'intelligenza artificiale.

Si pensi alla schermata della piattaforma *Netflix* che differisce da persona a persona in base alle preferenze di ciascuno. Ebbene, tali preferenze vengono individuate da un sistema di machine learning sulla base, tra gli altri fattori, dei film precedentemente visti dai quali vengono estrapolati innumerevoli dati: genere del film, attori protagonisti, carattere dei personaggi (un personaggio femminile dal carattere forte), genere della colonna sonora, ecc. Questo procedimento è risultato talmente accurato che nel 2013 Netflix è entrata nel settore della produzione presentando la sua prima serie televisiva "*House of*

---

<sup>54</sup> Questi valori sono: 1. Azione umana e supervisione 2. Robustezza tecnica e sicurezza 3. Privacy e Governance dei dati 4. Trasparenza 5. Diversità, non discriminazione ed Equità 6. Benessere Sociale ed Ambientale 7. Accountability

<sup>55</sup> Per un commento alla partita, G. Kasparov, *Deep thinking: Where Machine Intelligence Ends and Human Creativity Begins*, Hodder & Stoughton, London, 2017. Si veda anche F. H. Hsu, *IBM's Deep Blue Chess grandmaster chips*, in «IEEE Micro», 2, 1999, pp. 70-81.

*cards*” il cui successo era, già in partenza, assicurato, in quanto ogni dettaglio fu creato e modellato sulla base delle preferenze dei clienti<sup>56</sup>.

I prodotti dell’A.I. con cui siamo soliti entrare in contatto nella vita di tutti i giorni sono pertanto innumerevoli, e gli esempi potrebbero continuare all’infinito ma, non essendo questo lo specifico settore di analisi preso in considerazione, si ritiene sufficiente in tale sede menzionare brevemente che lo stesso algoritmo di machine learning viene inoltre utilizzato da *Siri*, l’assistente vocale di Apple, dai motori di ricerca utilizzati per navigare in rete, così come dall’applicazione “Waze” che, a differenza del predecessore “Tom Tom”, individua il percorso più rapido evitando così code, incidenti stradale e rallentamenti, apprendendo in maniera costante dalle abitudini degli utenti.

Gli esempi sopra esposti mettono in luce come un errore provocato da Waze o da Siri può causare al più qualche fastidio per l’utente. Tuttavia, l’intelligenza artificiale viene ad oggi impiegata anche in settori particolarmente delicati, come quello della sanità, della giustizia, nell’ambito bellico e in quello dei trasporti pubblici. In tali casi, un eventuale errore nel funzionamento risulterebbe idoneo a produrre un danno considerevole, facendo così sorgere il problema dell’attribuzione della responsabilità, anche penale.

Nonostante la trattazione prenda in esame esclusivamente il problema della responsabilità per i danni provocati dalle automobili driverless, si ritiene opportuno fare una breve ricognizione dell’impiego dell’A.I. anche negli altri settori citati. Le problematiche connesse all’attribuzione della responsabilità infatti sono in parte condivise per tutti i prodotti in grado di auto apprendere.

#### SETTORE MEDICO

Nell’ambito medico l’intelligenza artificiale gioca un ruolo sempre più fondamentale, in particolare modo nel settore della diagnostica, grazie anche a una sempre maggiore disponibilità di dati e informazioni.

La medicina moderna, e nello specifico il settore della diagnostica, si fonda sulla ricerca di elementi ricorrenti all’interno dei dati disponibili, *pattern*<sup>57</sup>. Si tratta, in particolare, di attività nelle quali gli algoritmi si dimostrano particolarmente efficienti. Non stupisce

---

<sup>56</sup> Per un approfondimento sul tema si veda H. Fry, *Hello World. Essere umani al tempo del digitale*, cit. pp. 171-190.

<sup>57</sup> Termine inglese traducibile a seconda dei casi come “modello, elemento ricorrente, struttura ripetitiva”. Il riconoscimento del pattern è una delle componenti essenziali del machine learning.

infatti che l'intelligenza artificiale sia già oggi in grado di identificare con grande precisione alcune malattie, come ad esempio i tumori, nelle immagini radiologiche grazie alla possibilità di avere accesso istantaneo a un patrimonio storico di dati non alla portata di un singolo medico il quale, per ragioni legate a stanchezza o inesperienza, può commettere errori in questa fase delicata<sup>58</sup>. Si cita a tal proposito la piattaforma sviluppata da IBM “*Watson for Oncology*”, sistema di intelligenza artificiale adottato in oltre 50 ospedali nel mondo come strumento di supporto alle decisioni cliniche e terapeutiche<sup>59</sup>.

La preoccupazione circa l'eventualità che il lavoro del medico possa essere in futuro soppiantato da una macchina tuttavia è ancora lontana. Sono infatti numerosi gli aspetti che un algoritmo non è oggi, e probabilmente non sarà mai, in grado di soddisfare: l'empatia ad esempio, o la capacità di aiutare i pazienti ad affrontare difficoltà sociali e psicologiche.

Uscendo dall'ambito della diagnostica risulta doveroso accennare agli enormi progressi dell'Istituto Sant'Anna di Pisa nel settore delle protesi bioniche, in grado di comunicare col cervello inviando e ricevendo impulsi. Protesi, in grado di far percepire nuovamente tutte le sensazioni perse a seguito dell'amputazione di un arto e in grado di auto apprendere e auto gestirsi grazie all'impiego delle reti neurali artificiali<sup>60</sup>.

Un ultimo riferimento, di grande attualità, attiene poi all'impiego dell'A.I. nella lotta al COVID-19, virus che ha messo, e sta tuttora mettendo, in ginocchio il pianeta scatenando una vera e propria pandemia.

Al fianco dei medici si sono schierati i colossi Alibaba, Baidu e Tencent che hanno messo a disposizione un vero e proprio arsenale di tecnologie: dal nuovo sistema di diagnosi che permette di rilevare, tramite scansioni tomografiche computerizzate (quindi tramite TAC), nuovi casi di coronavirus con un tasso di accuratezza fino al 96% e in un tempo di 20 secondi (abbattendo pertanto notevolmente i tempi d'attesa dei tradizionali tamponi), alle “telecamere intelligenti” in grado di intercettare le persone che non indossano una

---

<sup>58</sup> H. Fry, *Hello World. Essere umani al tempo del digitale*, cit., pp. 80 e ss.

<sup>59</sup> ASCO *Nuove correlazioni tra l'uso della tecnologia cognitiva di Watson e la cura del cancro*, in «IBM news Room», 1 giugno 2017.

<sup>60</sup> E. D'Anna, G. Valle, A. Mazzoni, I. Strauss, et al., *A closed-loop hand prosthesis with simultaneous intraneural tactile and position feedback*, in «Science Robotics», 27, 2019, pp. 1-11.

mascherina, ma anche di effettuare una scansione termica in tempo reale così da individuare eventuali casi di febbre<sup>61</sup>.

Inoltre, Tencent, la holding che sta dietro alla popolare app di messaggistica WeChat (la più diffusa in Cina), ha lanciato l'applicazione “*close contact detector*” che avvisa gli utenti se entrano in contatto con un potenziale cittadino portatore di virus, e il maggior operatore telefonico del Paese, China Mobile, ha condiviso con alcuni media i dati di spostamento dei suoi utenti affetti da virus: dal treno preso, fino alla metropolitana o al supermercato. E questo è servito a tracciare in determinate città le possibilità di contagio<sup>62</sup>.

#### SETTORE GIURIDICO

L’A.I. sta facendo il suo ingresso anche nell’ambito giuridico. Tra le principali modalità applicative dell’intelligenza artificiale risalta sicuramente la c.d “Giustizia predittiva”. Con tale termine si fa riferimento innanzitutto a quegli “*strumenti di supporto alla funzione legale, e solo poi giurisdizionale, capaci di analizzare in tempi brevi, più brevi di quelli concessi all’uomo, una grande quantità di informazioni con l’obiettivo di prevedere l’esito, o i possibili esiti, di un giudizio*”<sup>63</sup>.

La giustizia predittiva conferma e risponde a quell’innata tendenza dell’uomo alla certezza del diritto. Sosteneva Montesquieu, già nel 1777, l’idea secondo cui i giudici dovessero rappresentare non solamente la bocca della legge, ma soprattutto “*êtres inanimés*”<sup>64</sup>. L’espressione indicava che non è solamente il contenuto delle regole a fare la differenza tra un regime autarchico e un regime democratico, ma altresì il modo in cui

---

<sup>61</sup> B. Sim, *Coronavirus, l’intelligenza artificiale di Alibaba lo identifica in 20 secondi con una Tac*, in «Sole 24 ore», 6 marzo 2020.

<sup>62</sup> B. Simonetta, *Così big data e intelligenza artificiale stanno battendo il coronavirus in Cina*, in «Sole 24 ore», 9 marzo 2020.

<sup>63</sup> E. Rulli, *Giustizia predittiva, intelligenza artificiale e modelli probabilistici. Chi ha paura degli algoritmi?*, in «Analisi Giuridica dell’Economia», 2, 2018, p. 539. Si devono pertanto escludere dalla nozione tutti quegli strumenti utilizzati da magistrati e avvocati al solo scopo di razionalizzare l’impiego delle risorse: i motori di ricerca come elemento sostitutivo alla ricerca condotta manualmente, l’impiego di procedure telematiche in sostituzione dei depositi cartacei e infine l’utilizzo di software di correzione automatica impiegati al fine di ridurre errori grammaticali nella stesura degli atti.

<sup>64</sup> Tradotto “*esseri inanimati*”. Montesquieu, *Lo spirito delle leggi*, III ed., a cura di S.Cotta, UTET, Milano, 1996, 326-327.

queste vengono applicate. Gli “esseri inanimati”, così come i sistemi di giustizia predittiva, sarebbero pertanto privi di quell’emotività e di quella discrezionalità che connota gli esseri umani, garantendo in questo modo la certezza dell’applicazione della regola scritta.

L’intelligenza artificiale può essere quindi impiegata in ambito giuridico secondo diverse modalità: come strumento di prevenzione della criminalità<sup>65</sup>, come strumento integrativo dell’attività del giurista<sup>66</sup> e, infine, come strumento di previsione dell’esito di un giudizio. Si ricorda, a quest’ultimo proposito, il noto algoritmo statunitense *Compass*, che non entra nel merito del giudizio penale, ma piuttosto viene impiegato in America dall’organo giudicante per una valutazione dei rischi di recidiva dei condannati, facilitando così la decisione del giudice di pronunciarsi sull’opportunità o meno rilasciare un soggetto su cauzione<sup>67</sup>.

L’impiego massiccio di sistemi di intelligenza artificiale all’interno del sistema giustizia comporterà quindi profondi mutamenti nel modo stesso di intendere la professione legale. Sarà opportuno, se non necessario, ripensare il piano di studi universitari in modo tale che possa essere integrato con alcuni principi e nozioni basilari di matematica e di logica, “*funzionali a creare le basi per un apprendimento continuo in campo informatico*”<sup>68</sup>.

D’altronde, lo stesso proposito è stato manifestato anche in sede Europea, la quale, come sopra accennato, esorta gli Stati membri a considerare adeguatamente gli aspetti riguardanti l’istruzione, sottolineandone a tal proposito la necessità di includere in tutti i programmi di insegnamento, fin dai primi anni scolastici, le competenze digitali<sup>69</sup>.

---

<sup>65</sup> La funzione è quella di anticipare orari e luoghi in cui potrebbero essere commessi crimini della stessa specie a seguito dell’immissione, all’interno del computer, di una serie di dati estrapolati ad esempio da denunce. Si rileva a tal proposito il sistema informatico *X-LAW* in uso da parte di alcune questure italiane. Grazie al suo utilizzo sono stati raggiunti ottimi risultati nel campo della prevenzione della criminalità.

<sup>66</sup> Si tratta di sistemi utilizzati per l’interpretazione delle norme e per l’individuazione delle argomentazioni in grado di avvalorare la tesi che si intende sostenere.

<sup>67</sup> Oltre all’algoritmo *Compass*, in Francia è on line [predictice.com](http://predictice.com), una piattaforma in grado di anticipare il potenziale risultato di una controversia e prospettare la migliore strategia giudiziaria partendo dall’analisi dei dati.

<sup>68</sup> Così E. Rulli, *Giustizia predittiva, intelligenza artificiale e modelli probabilistici. Chi ha paura degli algoritmi?*, cit., p. 539.

<sup>69</sup> Parlamento Europeo, *Una politica industriale europea globale in materia di robotica e intelligenza artificiale*, Strasburgo, 12 febbraio 2019.

Risulterà pertanto necessario ampliare l'ambito delle competenze dei giuristi, si rischierebbe altrimenti di cadere nel paradosso prospettato dal romanzo del giuslavorista francese P. Janot<sup>70</sup>.

#### SETTORE BELLICO

Sorgono inoltre numerosi interrogativi etici sull'impiego, nel settore bellico, di armi letali autonome<sup>71</sup> e sui robo-soldati, tanto che nel 2017 venne lanciato un appello all'Onu da parte di 116 fondatori di aziende di robotica e intelligenza artificiale - tra cui Elon Musk, da sempre sensibile a questo tema, e Mustafa Suleyman, fondatore di DeepMind (Google) con cui si chiedeva all'organizzazione internazionale di *“trovare modi per prevenire una corsa agli armamenti autonomi, per proteggere i civili dagli abusi e per evitare gli effetti destabilizzanti di queste tecnologie”*<sup>72</sup>.

#### SETTORE TRASPORTI

L'intelligenza artificiale viene infine impiegata, a dosi sempre maggiori, nell'ambito dei trasporti, tanto che il Parlamento Europeo rileva come il settore dell'automobile sia quello in cui si avverte con maggiore urgenza la necessità di interventi normativi a livello comunitario (oltre che mondiale) allo scopo di garantire la commercializzazione e la circolazione transfrontaliera di veicoli a guida autonoma. Evidenziando inoltre che *“approcci normativi frammentari ostacolerebbero l'attuazione dei sistemi di trasporto autonomi e metterebbero a repentaglio la competitività europea”*<sup>73</sup>.

---

<sup>70</sup> P. Janot, *Lex Humanoïde, des robots et des juges*, Editions Thot, Fontaine, 2017. L'autore nel suo romanzo immagina un anno 2030 nel quale la giustizia è interamente amministrata da una piattaforma digitale, Cujas, i giudici sono sostituiti da algoritmi, e il ruolo degli avvocati è ridotto ad un'asettica compilazione di formulari.

Il sistema è apprezzato sia da cittadini che istituzioni in quanto consente di eliminare la soggettività tipica del giudizio umano, e solo successivamente si scopre che dal database di Cujas sono stati eliminati gli scritti di importanti giuristi. Nonostante il sistema venga sofferto e criticato da un giovane avvocato, frustrato dalla fredda meccanizzazione del proprio mestiere, l'unico soggetto in grado di scoprire la falla esistente nel sistema è stato un ingegnere informatico. Per evitare quindi che una falla della giustizia amministrata da sistemi dell'A.I siano comprensibili esclusivamente ai tecnici sarà necessario che anche, e soprattutto, chi un domani dovrà fronteggiare una *giustizia artificiale*, possa comprenderla a fondo.

La giustizia predittiva in definitiva potrà dunque rappresentare uno dei mezzi per decongestionare la macchina della giustizia specie se utilizzata per la definizione di piccole controversie. Per un approfondimento sul tema si veda H. Fry, *Hello World. Essere umani al tempo del digitale*, cit., pp. 54-79.

<sup>71</sup> Per un approfondimento sul tema: J. C. Rossi, *La guerra che verrà: le armi autonome*, in «IRIAD Review», 11, 2016, pp. 1-20. E M. Tegmark, *Essere umani nell'era dell'intelligenza artificiale*, Raffaello Cortina editore, Milano, 2018.

<sup>72</sup> S. Hawking, E. Musk, S. Wozniak, N. Chomsky, et al., *An open letter to the united nations convention on certain conventional weapons*, 24th International Joint Conference on Artificial Intelligence, Buenos Aires, 2015.

<sup>73</sup> Parlamento europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit., pt. nr. 25.

### 3. AUTO A GUIDA AUTONOMA

#### 3.1 Auto a guida autonoma come simbolo della IV rivoluzione

*“If I didn’t know better, I’d say a ghost was driving.”*

Così, il giornalista di Forbes, Joann Muller, descrive la sua esperienza a bordo della Google’s driverless car<sup>74</sup>.

Tra le diverse e più affascinanti applicazioni dell’A.I spicca in particolare il tema delle auto a guida autonoma, realtà sempre più vicina e concreta.

Quale simbolo della rivoluzione industriale del XIX secolo, l’auto è probabilmente vicina ad una nuova rivoluzione, quella industriale e tecnologica. Il concetto di veicolo cambierà sicuramente passando da un semplice strumento di mobilità ad un nuovo stile di vita. L’avvento del veicolo autonomo promette anche profondi cambiamenti in molti altri settori, come quelli assicurativi e legislativi, facendo così sorgere innumerevoli quesiti anche etici<sup>75</sup>: sarà possibile punire l’uomo per un errore commesso da una macchina? Chi sarà quindi responsabile in caso di incidente? In una situazione estrema di scelta, il software chi dovrà salvare, la folla o il passeggero, il bambino o la donna incinta?

Al fine di affrontare nella maniera più completa possibile tali interrogativi risulta tuttavia necessaria una breve ricognizione di ciò che rappresentano le self driving car e dei livelli di automazione raggiunti ad oggi.

Un’autovettura autonoma è *“un veicolo automatico in grado di soddisfare le principali capacità di trasporto di un’automobile tradizionale”*<sup>76</sup>.

È, nello specifico, un veicolo che utilizza una combinazione di sensori, telecamere, radar e intelligenza artificiale per spostarsi tra diverse destinazioni senza necessità di un intervento umano, su strade che non siano state pre adattate.

In realtà, automobili che rispondono perfettamente a tale definizione, ancora non sono presenti nelle nostre strade, se non in fase sperimentale.

---

<sup>74</sup>J. Muller, *No Hands, No Feet: My Unnerving Ride In Google’s Driverless Car*, in «Forbes», 21 marzo 2013.

<sup>75</sup> Parlamento europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit., pt. nr. 27.

<sup>76</sup> Definizione fornita da M. C. Gaeta, *Automazione e responsabilità civile automobilistica*, «in Responsabilità civile e previdenza», 5, 2016, pp. 1718.

Pertanto, attualmente il termine, che dovrebbe essere riferito esclusivamente a quelle automobili interamente automatizzate, viene utilizzato, anche nei più importanti documenti istituzionali, per indicare livelli di automazione minori<sup>77</sup>.

La SAE international<sup>78</sup> a tal proposito ha individuato, codificandoli in questa maniera a livello internazionale, 5 diversi livelli di automazione attraverso una scala che va da 0 a 5. La classificazione ad opera della SAE international<sup>79</sup> parte dunque da un Livello 0 di automazione, utilizzato per indicare quelle automobili tradizionali in cui è il guidatore a doversi occupare, senza supporti elettronici, di ogni aspetto relativo alla guida, fino ad un Livello 5, nel quale il veicolo è in grado, senza alcun tipo di intervento umano, di gestire in piena autonomia la guida su qualsiasi tipo di percorso<sup>80</sup>.

---

<sup>77</sup> Tra questi si ricorda la Risoluzione Parlamento europeo del 16/2/2017 recante “*Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*”, ove, nella voce “Mezzi di trasporto autonomi” pt. nr. 24, sottolinea che “*la nozione di trasporto autonomo include tutte le forme a pilotaggio remoto, automatizzate, connesse, e autonome di trasporto*”. Tra i documenti più recenti si menziona anche la Risoluzione del Parlamento Europeo del 12/2/2019 su “*Una politica industriale europea globale in materia di robotica e intelligenza artificiale*” nel paragrafo 3.1.4, pt. nr. 89. A livello nazionale il termine è stato utilizzato anche nel documento della Camera dei Deputati, servizio studi-Dipartimento dei trasporti, *La mobilità e il futuro: l'auto a guida autonoma*, Dossier n° 275, 31 gennaio 2017.

<sup>78</sup>SAE International (SAE) Society of Automotive Engineers, è un ente di normazione nel campo dell'industria aerospaziale, automobilistica e veicolistica

<sup>79</sup> SAE International, *Level of driving automation are defined in new Sae international standards J3016*, in [sae.org](http://sae.org), 2014.

<sup>80</sup>Per completezza espositiva si riportano qui tutti i livelli indicati dal documento SAE International's new standard J3016: Livello 0: nessuna automazione. Il guidatore ha il completo controllo del veicolo e del monitoraggio della strada; Livello 1: il veicolo è dotato di una o più funzioni di controllo che operano indipendentemente l'una dall'altra; il guidatore ha il controllo generale del veicolo ma può attivare alcune funzioni di sicurezza, come il cruise control adattivo, i controlli elettronici di stabilità del veicolo (ESC), il supporto dinamico di frenata, i sistemi di mantenimento della corsia di marcia (lane centering), che operano comunque sotto il suo controllo e lo aiutano nella guida, ma che non operano congiuntamente su più funzioni e quindi non assumono il controllo del veicolo al suo posto; tale livello di automazione implica sempre un controllo continuo e attivo sul volante e sui pedali da parte del guidatore, che non possono essere contemporaneamente abbandonati; Livello 2: il veicolo usufruisce di una combinazione di almeno due funzioni primarie automatiche di controllo del veicolo che agiscono congiuntamente e consentono al guidatore di cedere il controllo di alcune funzioni in situazioni limitate. Il guidatore è obbligato comunque a monitorare la strada ed essere in grado di intervenire immediatamente, anche in caso di attivazione delle funzioni autonome da parte del veicolo (ad es. il cruise control adattivo combinato con il lane centering, il sistema di mantenimento della corsia); a tale livello di automazione il guidatore, quando il sistema è attivo, può togliere momentaneamente le mani dal volante o i piedi dai pedali; Livello 3: è la guida autonoma limitata (Limited self-driving automation) in cui il veicolo consente al guidatore di cedere al sistema di guida autonoma il pieno controllo delle funzioni critiche di sicurezza, ma limitatamente a precise condizioni di traffico e ambientali. Il guidatore è comunque tenuto ad un controllo, anche se non continuo, di tali funzioni ma il sistema è in grado di assicurare il controllo del veicolo e di segnalare l'esigenza di tornare alla modalità di controllo manuale del veicolo; Livello 4: è la guida autonoma piena (Full self-driving automation), in cui il veicolo può svolgere tutte le funzioni critiche di sicurezza e monitorare costantemente le condizioni della strada per l'intero viaggio, ma il sistema non è in grado di operare in qualsiasi condizione di viaggio, esempio in condizioni estreme di tempo. Nelle condizioni in cui il sistema può operare il guidatore fornirà gli input di navigazione autonoma e non dovrà controllare il veicolo costantemente durante il viaggio; Livello 5 di automazione: è richiesta solo l'indicazione della destinazione e l'avvio del sistema, senza altro intervento da parte del guidatore, in qualsiasi condizione in cui possa guidare un conducente. Definizioni utilizzate da Camera dei Deputati, Servizi Studi, *La mobilità del futuro: l'auto a guida autonoma*, in Dossier n° 275 - Schede di lettura, 31 gennaio 2017.

Ed è solo a veicoli dotati di tali caratteristiche, peraltro non ancora in commercio<sup>81</sup>, che bisognerebbe formalmente attribuire la denominazione di automobili driverless.

### 3.2 Quali Vantaggi

Il mercato dell'auto a guida autonoma è dotato di enormi potenzialità non solo sul piano economico ma anche sul piano sociale.

Se l'intelligenza artificiale è il motore della nuova rivoluzione industriale, l'automobile, mezzo di cui è difficile fare a meno al giorno d'oggi, non potrà non esserne travolta, e sarà destinata, almeno nel lungo periodo, a diventare interamente automatizzata.

Gli studiosi hanno colto a pieno tali potenzialità e hanno infatti previsto che il mercato dell'auto self driving crescerà esponenzialmente nei prossimi anni.

Secondo uno studio del 2018 pubblicato sull'hub scientifico dell'UE, si prevede che la piena diffusione delle tecnologie dei veicoli connessi e automatizzati (CAV) avrà un impatto sostanziale sull'economia europea. Il mercato dei veicoli a guida autonoma crescerà esponenzialmente, portando così alla creazione di nuovi posti di lavoro e al raggiungimento, entro il 2025, di profitti pari a €620 miliardi per il settore automotive e a €180 miliardi per quello elettronico<sup>82</sup>.

Strategy Analytics, a tal proposito, ha realizzato uno studio nel 2017 per conto di Intel che mostra come il mercato dei veicoli autonomi avrà un valore globale che passerà dagli \$800 miliardi del 2035 (anno di riferimento dello studio) ai \$7 trilioni del 2050<sup>83</sup>.

Date le enormi potenzialità presentate a livello economico, la prospettiva dello sviluppo delle auto a guida autonoma è, ed è stato, oggetto di ampio studio, sia da parte di istituzioni operanti a livello internazionale ed europeo, che da parte di numerosi centri di ricerca operanti nel settore dei trasporti. Da tali studi sono emerse alcune questioni comuni

---

<sup>81</sup> Secondo quanto riferito dal sito del Parlamento Europeo è attualmente consentita in Europa la commercializzazione dei veicoli a guida assistita di livello 1 e 2, mentre si prevede che le auto a guida autonoma di livello 3 e 4, attualmente in fase di collaudo e verifica, faranno il loro ingresso sul mercato nel decennio compreso tra il 2020 e 2030. Si dovrà infine, sempre secondo le aspettative del Parlamento Europeo, attendere il 2030 per i veicoli a guida interamente autonoma di livello 5. Cfr. Parlamento Europeo, *Auto a guida autonoma in UE: dalla fantascienza alla realtà*, in [europarl.europa.eu](http://europarl.europa.eu), 14 gennaio 2019.

<sup>82</sup> Parlamento Europeo, *Auto a guida autonoma in UE: dalla fantascienza alla realtà*, cit.

<sup>83</sup> R. Lanctot, H. Cohen, *Accelerating the Future: The Economic Impact of the Emerging Passenger Economy*, in «Strategy Analytics», 1, 2017, pp.11.

che rappresentano, tra l'altro, anche i principali obiettivi che questa nuova forma di mobilità consentirà di raggiungere:

- Una maggiore sicurezza negli spostamenti, e quindi la riduzione degli incidenti. L'organizzazione mondiale della sanità stima che gli incidenti stradali rappresentano ogni anno la morte di 1,35 milioni di persone e il numero dei feriti è stimato a oltre 50 milioni in tutto il mondo<sup>84</sup>. Si consideri inoltre che la maggior parte degli incidenti stradali è dovuto a errori umani quali stanchezza, distrazione ed elevato tasso alcolemico, e, il sistema di guida autonoma risulterebbe in grado di apportare una riduzione di tali eventi disastrosi del 90% salvando così milioni di vite l'anno<sup>85</sup>. Le macchine non si stancano e non si ubriacano e sono inoltre meno soggette ai rischi legati alle condizioni ambientali, quali mancanza di luce o presenza di nebbia. Gli ostacoli maggiori sicuramente attengono alla difficoltà nella progettazione di tali sistemi, ma una volta superate completamente, tale scenario potrà diventare realtà. Ovviamente il rischio zero non esiste e le questioni etiche che si prospettano dinanzi a scenari che impongono di effettuare una scelta drastica in caso di incidenti imminenti e imprevedibili permangono.

- La possibilità di muoversi più facilmente per anziani e disabili
- La riduzione del traffico e dell'inquinamento. Grazie alle smart road il traffico dovrebbe essere più fluido con conseguente riduzione delle emissioni inquinanti. Le auto automatizzate e connesse, nella scelta del percorso, prediligeranno la strada senza rallentamenti, incidenti o lavori in corso consentendo così una miglior redistribuzione del traffico cittadino. Tale previsione è in realtà controversa, molti esperti, tra cui si cita Lew Fulton, direttore dell'Institute of Transportation Studies dell'Università della California di Davis, affermano, al contrario, che l'impatto sul traffico di tali veicoli sarà minimo o addirittura negativo, almeno in assenza di una tassazione che scoraggi la circolazione senza persone a bordo. Si parla dell'eventualità che le aziende pro-

---

<sup>84</sup> World Health Organization, *Global status report on road safety*, 2018

<sup>85</sup> AON, *Global Analytics*, 2016.

grammino le auto a guida autonoma affinché facciano commissioni o ritirino pacchi aumentando in questo modo il traffico e moltiplicando la circolazione di automobili prive di esseri umani, le “auto zombie”, ha esemplificato Fulton.<sup>86</sup>

- Aumento del car-sharing con conseguente riduzione della spesa familiare per l’acquisto e la manutenzione delle auto. Toyota sta effettuando investimenti per oltre 500 milioni di \$ in Uber finalizzati a sviluppare un servizio di ride sharing basato sul van Toyota Sienna a guida autonoma<sup>87</sup>.
- Soluzione durante le pandemie? L’esperienza Cinese insegna come l’automobile driverless sia una soluzione da esplorare in caso di emergenze, come quella generata dal Coronavirus, che obbligano le persone ad evitare il più possibile i contatti. Le misure di contenimento della pandemia hanno portato i consumatori a riflettere sull’opportunità di far consegnare merci a sistemi robotici che, ovviamente, non possono contagiare né restare contagiati. Neolix, un produttore cinese di camioncini per le consegne urbane totalmente robotizzati, ha visto incrementare notevolmente i suoi ordini negli ultimi tempi. Ha infatti schierato numerosi veicoli, alti poco più di un metro, in 10 grandi città in Cina per effettuare le consegne oltre che per svolgere il servizio di disinfezione.

### **3.3 Ostacoli e problemi aperti.**

Nonostante le enormi potenzialità che caratterizzano le self driving car, ad oggi, rimangono ancora aperte diverse questioni che ne rallentano lo sviluppo e che dovranno necessariamente essere risolte prima di poter veder circolare autonomamente i veicoli per le strade di tutto il mondo.

#### ACCETTAZIONE DEI CONSUMATORI

Gli esperti del settore e le case automobilistiche concordano sul fatto che i veicoli autonomi domineranno la scena nel prossimo futuro, affinché lo scenario da questi prospettato diventi realtà, sarà tuttavia necessario che il pubblico dei consumatori, ai quali il prodotto

---

<sup>86</sup> Questo quanto contenuto in una dichiarazione di Lew Fulton al Boston Globe nel 2017. A tal proposito si ricorda che gli amministratori del Massachusetts, USA, hanno già proposto nello stesso anno una tassa sulle auto a guida autonoma al fine di prevenire una situazione del genere.

<sup>87</sup> Il lancio è atteso per il 2021. “L’accordo è il primo di questo genere e segna l’impegno nel portare tecnologie di livello mondiale nella rete Uber”, ha dichiarato l’amministratore delegato di Uber, Dara Khosrowshahi. D. La Cava, *Toyota investe 500 mln\$ in Uber, sinergie nelle auto a guida autonoma*, in «Finanza Online», 28/8/2018.

è rivolto, accetti con entusiasmo il cambiamento apportato da questa nuova forma di mobilità.

Come si analizzerà in seguito<sup>88</sup>, la problematica non riveste esclusivamente carattere teorico. Ad oggi infatti, un prodotto, per essere commercializzato deve rivestire le caratteristiche di un prodotto sicuro<sup>89</sup>, ossia un prodotto che, “*in condizioni di uso normali o ragionevolmente prevedibili, presenti esclusivamente rischi minimi*”<sup>90</sup>.

Tuttavia, e specie nei settori fortemente innovativi, nel calcolo del rischio risulterà necessario tenere in considerazione, non esclusivamente la sua oggettiva consistenza, ma prima ancora, la sua percezione da parte della società, rappresentata dai consumatori stessi<sup>91</sup>.

Così prima che le questioni etico-giuridiche connesse all’attribuzione della responsabilità possano essere trattate, dovrà essere affrontata tale problematica. In altre parole, le questioni connesse alla responsabilità avranno seguito solamente nel caso in cui tali prodotti vengano commercializzati, e ciò dipenderà in parte anche dalla risposta che la società avrà nei confronti degli stessi.

Sono proprio i timori della collettività nei confronti di questa nuova forma di tecnologia che potrebbero rappresentare uno dei maggiori ostacoli alla guida interamente automatizzata.

Secondo uno studio effettuato dal MIT, *Advanced Vehicle Technology*, i principali motivi della mancata approvazione verso un mezzo completamente autonomo sono rappresentati dalla volontà di mantenere il controllo, dalla mancanza di fiducia nella tecnologia e infine dal timore che un’auto del genere non funzionerà mai correttamente<sup>92</sup>.

Ciò tuttavia non deve scoraggiare, infatti, citando le parole di Deming William Edwards,<sup>93</sup> “*le due principali regole che stanno alla base della vita stessa sono:*

*1. Il cambiamento è inevitabile.*

---

<sup>88</sup> Cfr. cap. 2, par. 2.2 di questo lavoro, relativo alla *Società del rischio*, pp. 49 e ss.

<sup>89</sup> *Codice del consumo*, Dlgs n°206 del 06/09/2005, Art 104, c 1.

<sup>90</sup> Codice del consumo articolo 103 c1 lett a). Si riporta qui l’intero articolo 103 c1 lett a) “*Ai fini del presente titolo si intende per prodotto sicuro: qualsiasi prodotto, come definito all’articolo 3, comma 1, lettera e), che, in condizioni di uso normali o ragionevolmente prevedibili, compresa la durata e, se del caso, la messa in servizio, l’installazione e la manutenzione, non presenti alcun rischio oppure presenti unicamente rischi minimi, compatibili con l’impiego del prodotto e considerati accettabili nell’osservanza di un livello elevato di tutela della salute e della sicurezza delle persone.*”

<sup>91</sup> Così, C. Piergallini, *Danno da prodotto e responsabilità penale, profili dogmatici e politico-criminali*, Giuffrè editore, Milano, 2004, pp. 17 e ss.

<sup>92</sup> L. Fridman, B. Reimer, L. Angell, S. Seaman et al., *MIT Advanced Vehicle Technology Study: Large-Scale Naturalistic Driving Study of Driver Behavior and Interaction with Automation*, in «IEEE Access», 11, 2019, pp. 102031.

<sup>93</sup> Ingegnere, saggista, docente e consulente di gestione aziendale e manager statunitense a cui fu riconosciuto il merito per gli studi sul “*miglioramento della produzione*” negli Stati Uniti d’America durante la seconda guerra mondiale.

## 2. *Tutti cercano di resistere al cambiamento.*”

La società si presenta spesso restia alla novità e necessita del tempo prima di accettarla completamente, soprattutto nel momento in cui questa è in grado di mettere in discussione le certezze che hanno sempre accompagnato la nostra vita.

I numeri tuttavia risultano in parte incoraggianti, dai diversi studi effettuati risulta infatti in crescita il numero delle persone che si dichiarano favorevoli all'introduzione dei veicoli autonomi.

Cap Gemini, ad esempio, in un'analisi effettuata nel 2017<sup>94</sup> sottoponendo 8000 persone di 8 diversi mercati a un sondaggio sulla guida autonoma, ha confermato un trend positivo nell'accettazione di tale tecnologia da parte dei consumatori. In tutti i paesi oggetto dell'analisi è infatti drasticamente calata, rispetto i due anni precedenti, la percentuale dei consumatori che riservano dubbi sulla sicurezza delle auto driverless. Lo stesso trend è stato confermato in uno studio condotto da Deloitte l'anno successivo, rilevando una sempre maggior accettazione da parte degli automobilisti rispetto l'anno precedente. In Italia, ad esempio, si è passati dal 63% di consensi del 2017 al 70% del 2018, in un solo anno quindi il 7% degli intervistati ha mutato il proprio giudizio nei confronti di tali veicoli<sup>95</sup>.

É comunque provando la guida autonoma che si imparerà a conoscerla ed accettarla. In questo senso, le case automobilistiche hanno effettuato una corretta manovra di mercato introducendo gradualmente le funzioni di assistenza alla guida piuttosto che presentare, in un futuro più o meno prossimo, un'autovettura completamente automatizzata. Questo permette, da un lato alla casa produttrice di ricevere feedback e correggere eventuali difetti, dall'altro consente al cliente di adattarsi più facilmente al cambiamento. L'automobilista odierno è infatti abituato a strumenti di ausilio alla guida, ad esempio l'Adaptive Cruise Control e l'Automatic Emergency Braking, che controllano rispettivamente l'adattamento della velocità alle diverse situazioni del traffico e la frenata in caso d'emergenza.<sup>96</sup>

---

<sup>94</sup> D. Godau, K. Grambow, N. Gill, R. Mehl, M. Winkler, *Beyond the Car*, CapGemini, 2017, pp. 28-29.

<sup>95</sup> C. Giffi, J. Vitale, *Global Automotive Consumer Study*, Deloitte, 2018, p. 2.

<sup>96</sup> Corporale vehicle observatory, *Le auto a guida autonoma: siamo già nel futuro?*, ottobre 2016, p. 7. A tal proposito la SAE International, sfruttando l'assioma per cui è insito nella natura umana temere ciò che non si conosce, ha lanciato, nell'Aprile 2019, l'idea di offrire un breve viaggio di 6 minuti a tutti coloro che si sono iscritti al programma di test con la Chrysler Pacifica Hybrid di Waymo-Google. Le recensioni riportate sono state sicuramente positive e l'esperienza è stata in grado di far ricredere molte persone sull'argomento, come è successo a un pensionato di 73 anni, Mary Van Der Maas che, a seguito dell'esperienza ha riportato *“Quando senti parlare di pericoli capisci subito che c'è tanta ignoranza sul tema”*.

Si accenna infine a un'esperienza recente già menzionata nel paragrafo precedente. Si parla dei numerosi veicoli automatizzati schierati dal produttore cinese Neolix per effettuare consegne e disinfettare le strade durante il periodo di quarantena imposto a causa dell'emergenza COVID-19.

Tale esperienza, come ha riportato Tallis Liu, manager del Business Development, alla testata IEEE Spectrum, ha condotto non solo a una maggiore accettazione di questo nuovo tipo di tecnologia ma altresì a un aumento della domanda da parte del pubblico in generale e delle istituzioni governative. Il manager cinese sottolinea infine come questo diverso atteggiamento nei confronti della guida autonoma rifletta i cambiamenti di comportamento che si verificano durante le grandi crisi.<sup>97</sup>

## QUESTIONI ETICHE

Una volta superato l'ostacolo relativo all'accettazione da parte del pubblico, sarà possibile affrontare le ulteriori questioni etiche e giuridiche.

Una delle tematiche più controverse legate alla guida autonoma è quella che riguarda la questione etica. A dimostrazione di ciò basti pensare al fatto che, ad occuparsi delle nuove tecnologie, e in particolare di quelle legate alla mobilità, non sono solo gli esperti di intelligenza artificiale e robotica, ma anche i filosofi.

La questione etica attualmente più discussa riguarda le modalità secondo le quali dovrebbe essere impostato il software nel caso in cui si prospetti, in sede di programmazione, una situazione di pericolo estremo ed imminente in cui ogni possibile scelta adottabile comporti inevitabilmente il sacrificio di una o più vite a discapito di altre: chi e cosa dovrebbe salvare?

La questione nasce dal fatto che, in una situazione in cui è l'uomo al volante, la scelta che egli assumerà sarà dettata dall'istinto, dalla paura: non è una scelta del tutto libera e consapevole. Al contrario, le decisioni di una vettura autonoma sono prese da un software, le cui modalità di azione sono state, in alcuni casi, precedentemente impostate a "mente fredda", risultando pertanto prive di quell'emotività tipica che connota l'essere umano nelle situazioni di estremo pericolo.

Ed è proprio la mancanza di istintività ed emotività nelle scelte che rappresenta il nodo centrale del problema.

---

<sup>97</sup> P. E. Ross, *Is Coronavirus Speeding the Adoption of Driverless Technology?*, in «IEEE Spectrum», 17.03.2020.

Se il software fosse impostato nel senso di sacrificare, in nome del benessere aggregato della popolazione, sempre e comunque il minor numero di persone a discapito quindi della vita dell'unico passeggero dell'auto driverless, (sposando pertanto tesi di natura consequenzialistica o utilitaristica) si giungerebbe a un disincentivo all'acquisto di questo genere di vetture. Al contrario, se le scelte fossero implementate nel senso di assicurare in ogni situazione l'incolumità del passeggero, ad essere messa in pericolo sarebbe piuttosto la sicurezza generale e il benessere aggregato.

Ovviamente situazioni di questo genere accadono regolarmente: ogni anno milioni di persone muoiono a causa di incidenti stradali, la preoccupazione però, in questo caso, nasce dalla presenza di un algoritmo che a priori sia in grado di determinare il soggetto da sacrificare.

Sulla tematica si scontrano diverse correnti filosofiche e alcuni paesi hanno già provveduto ad adottare codici etici, la cui accettazione costituisce una delle condizioni fondamentali per poter dare avvio alla fase di sperimentazione sulle strade pubbliche di quella nazione.

Il tema verrà in ogni caso approfondito nel terzo capitolo.

#### QUESTIONI GIURIDICHE

Un ultimo aspetto problematico da affrontare attiene infine al piano della responsabilità, sia civile che penale, per gli eventi lesivi cagionati da tali agenti.

Si tratta di una tematica particolarmente cara al Parlamento Europeo, che infatti sottolinea come *“il settore dell'automobile sia quello in cui si avverte con maggiore urgenza la necessità di interventi normativi a livello comunitario (oltre che mondiale) allo scopo di garantire la commercializzazione e la circolazione transfrontaliera di veicoli a guida autonoma.”*<sup>98</sup> evidenziando poi come *“approcci normativi frammentari ostacolerebbero l'attuazione dei sistemi di trasporto autonomi e metterebbero a repentaglio la competitività europea”*<sup>99</sup>. L'assenza di un indirizzo legislativo unitario volto a regolare il problema della responsabilità in tali scenari, potrebbe infatti rappresentare un grosso ostacolo allo

---

<sup>98</sup>Parlamento Europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit. pt. nr. 25.

<sup>99</sup> *Ivi*, pt. nr. 2.

sviluppo di tali tecnologie, con effetti negativi, non solamente in termini di progresso tecnologico, ma soprattutto in termini di competitività economica.

Le problematiche che hanno ostacolato l'introduzione di una normativa in materia sono numerose, e verranno dettagliatamente esaminate nei successivi capitoli. Per il momento ci si limita ad anticipare che, l'avvento di tale nuova tecnologia sarà inevitabilmente destinato a stravolgere gli ordinari criteri in base ai quali, ad oggi, viene imputata la responsabilità nel caso in cui si verifichi un sinistro stradale. Il prodotto dell'A.I., per le particolari modalità tramite cui è in grado di apprendere dall'esperienza e agire autonomamente, non risulta infatti equiparabile agli ordinari prodotti oggetti, ma allo stesso tempo il suo livello di autonomia non consente di qualificarlo alla stregua di un vero e proprio agente. Scomparsa pertanto la figura del conducente ad assorbire su di se le colpe per gli eventi lesivi verificatisi, la dottrina sta da tempo individuando diversi soggetti potenzialmente idonei ad assumere la responsabilità per i danni provocati dagli agenti artificiali, quali il programmatore del software, i produttori o addirittura lo stesso robot. Si tratta tuttavia di proposte e teorie spesso tra loro contrastanti e soprattutto orientate verso uno stravolgimento degli ordinari canoni e dei principi cardine del diritto penale classico. La tematica verrà comunque esaminata a fondo nel terzo capitolo.

### **3.4 Stato di evoluzione attuale/soggetti coinvolti?**

I soggetti coinvolti dal mercato delle auto a guida autonoma sono molteplici. Nell'era della quarta rivoluzione industriale, come già accennato precedentemente, i mercati per rimanere competitivi, devono necessariamente adattarsi alle mutate condizioni di mercato al fine di mantenere la propria posizione<sup>100</sup>.

Tale situazione è particolarmente evidente nel settore delle auto driverless, ma più in generale in tutti i settori dell'intelligenza artificiale, ove gli operatori economici coinvolti si sono dovuti confrontare con un problema non da poco: la produzione di un oggetto e la scrittura di un algoritmo richiedono competenze diverse. Chi sa costruire auto non sa scrivere software, chi sa scrivere software non sa costruire auto.

---

<sup>100</sup> Del tema si è già discusso nel cap. 1, par. 2.4 di questo lavoro, relativo alla *Quarta rivoluzione industriale*.

È per questo motivo che nel settore dell'intelligenza artificiale gli accordi tra aziende automobilistiche e startup specializzate nella scrittura di algoritmi di machine learning sono all'ordine del giorno<sup>101</sup>.

Così, ad esempio, Amazon ha investito 530 milioni di dollari nei primi mesi del 2019 nella startup della Silicon Valley Aurora<sup>102</sup> e nella stessa direzione si sta muovendo, dal 2018, il gruppo automobilistico Hyundai-Kia<sup>103</sup>. Sempre con la start-up Aurora poi è stato annunciato, nel Giugno 2019, l'accordo con FCA: un memorandum d'intesa con il quale FCA si impegna a integrare il sistema di guida autonoma *Aurora Drive* nei furgoni Fiat<sup>104</sup>. Tali sviluppi mettono in luce le possibili ricadute sul piano della responsabilità nel caso in cui un prodotto dell'intelligenza artificiale provochi un danno a terzi. La tematica verrà affrontata nello specifico nel corso della trattazione, per ora è sufficiente far presente che, in un simile contesto, il numero dei soggetti potenzialmente responsabili aumenta.

Nel caso in cui si opti per l'attribuzione della responsabilità a colui che ha costruito il prodotto dell'intelligenza artificiale<sup>105</sup>, sarà necessario ricercare la causa effettiva del danno, al fine così di risalire al soggetto realmente responsabile.

Si dovrà quindi distinguere la figura del produttore da quella del programmatore. Il primo è colui che sviluppa in concreto il prodotto per la produzione e la diffusione di massa, mentre il programmatore è colui che crea gli algoritmi e i programmi attraverso cui tali prodotti agiscono.

Introducendo brevemente ciò che verrà approfondito in seguito, in sede comunitaria<sup>106</sup>, così come anche nella dottrina maggioritaria<sup>107</sup> l'idea prevalente è quella secondo cui, al fine di regolamentare la responsabilità derivante dai danni provocati dal prodotto robo-

---

<sup>101</sup> Accordi che si compendiano in ingenti investimenti delle prime nelle seconde o addirittura con vere e proprie acquisizioni.

<sup>102</sup> La redazione di Forbes, *La tua prima auto a guida autonoma potrebbe essere di Amazon*, in «Forbes», 10 Febbraio 2019.

<sup>103</sup> S. Pini, *Guida autonoma, Hyundai e Kia entrano in Aurora*, in «Il Sole 24ore», 15 giugno 2019.

<sup>104</sup> G. Coppola, *Fiat Teams With Silicon Valley's Aurora on Self-Driving Tech*, in «Bloomberg Technology», 10 giugno 2019.

<sup>105</sup> Si è scritto “nel caso in cui” in quanto la questione deve ancora essere trattata e risolta in sede legislativa. Nella *Risoluzione Parlamento europeo del 16 febbraio 2017 recante Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica* sono state avanzate diverse proposte per la regolamentazione della responsabilità, tra le quali, tra l'altro, anche la creazione della “*personalità elettronica*”.

<sup>106</sup> Così emerge dalla *Risoluzione Parlamento europeo del 16 febbraio 2017 recante “Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit.

<sup>107</sup> Tra questi ad esempio U. Ruffolo, *Intelligenza artificiale e responsabilità*, Giuffrè Editore, Milano 2017, pp. 81 e ss.

tico, il punto di partenza sia rappresentato dalla normativa contenuta nel Codice del consumo<sup>108</sup>. Sebbene questa sia rivolta a regolare la sola responsabilità civile per i danni da prodotto difettoso, appare idonea a costituire una valida base di ragionamento anche per le riflessioni in ambito penale.

Nell'articolo 114 si stabilisce che *“Il produttore è responsabile del danno cagionato da difetti del suo prodotto”*<sup>109</sup> ove per produttore si fa riferimento *“al fabbricante del prodotto finito o di una sua componente, o al produttore della materia prima”*<sup>110</sup>.

È discusso infatti se, nel caso in cui il danno sia imputabile a un difetto dell'algoritmo, la figura del programmatore possa essere assimilata a quella del produttore. Ossia se la scrittura di un algoritmo possa essere equiparata alla *fabbricazione di un prodotto*.<sup>111</sup>

Si è sostenuto che nella nozione di prodotto, ai sensi del Codice del Consumo siano compresi anche i beni immateriali, di riflesso anche software o algoritmi<sup>112</sup>.

Il problema non risiede tuttavia nella sola assimilazione di un algoritmo al prodotto, quanto piuttosto nell'equiparazione di un prodotto di machine learning<sup>113</sup>, in grado di determinare la macchina a compiere scelte non direttamente ed integralmente prevedibili, al prodotto dannoso così come preso in considerazione dal Dlgs 206/2005.

Sono tutte questioni aperte, evidenziate anche in sede comunitaria<sup>114</sup>, che mettono luce sulla necessità di adattare al più presto la normativa ai nuovi sviluppi futuri.

---

<sup>108</sup>Dlgs n°206 del 6/9/2005, così come modificato dalla L. nr 37 del 3/5/2019 e dalla L. nr 31 del 12/4/2019.

<sup>109</sup> Dlgs 206/2005, Titolo II, Responsabilità per danno da prodotti difettosi, articolo 114, Responsabilità del produttore.

<sup>110</sup>Dlgs 206/2005, articolo 115, Prodotto e produttore. Si riporta di seguito l'articolo completo così modificato dal decreto legislativo 23 ottobre 2007, n. 221. *“Prodotto, ai fini del presente titolo, e' ogni bene mobile, anche se incorporato in altro bene mobile o immobile. Si considera prodotto anche l'elettricità'. Produttore, ai fini del presente titolo, e' 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 dell'allevamento, della pesca e della caccia, rispettivamente l'agricoltore, l'allevatore, il pescatore ed il cacciatore”*.

<sup>111</sup> Per riprendere le parole utilizzate nell'art 115 c2 bis, Codice del Consumo.

<sup>112</sup> In questo senso si esprime V. Cuffaro, *Codice del consumo*, Giuffrè Editore, Milano, 2012, p. 575. La teoria è stata in realtà criticata sul presupposto secondo cui, dal momento che l'elettricità, bene immateriale, è esplicitamente inserita nel novero dei prodotti di cui all'articolo 115 c2, le altre cose immateriali sarebbero al contrario escluse dall'ambito di applicazione della normativa. Così D. Wuyts, *The Product Liability Directive – More than two decades of defective products in Europe*, in «Journal European Tort Law», 5, 2014, pp. 1-34 ; o anche da K. Alhelt, *The Applicability of the EU Product Liability Directive to Software, Comparative and International Law*, in «Journal South Africa», 2, 2001, p. 200.

<sup>113</sup> Per prodotto di machine learning si intende il connubio tra la parte fisica e materiale del prodotto e la componente immateriale, rappresentata dall'algoritmo.

<sup>114</sup> In particolare modo dal Parlamento Europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit.

## **Capitolo 2**

### **Cause da cui sorge la responsabilità.**

# 1 INTRODUZIONE

## 1.1. Il nucleo della questione: il *Responsability Gap*

Come analizzato sopra, si stima che le auto a guida autonoma porteranno ad una riduzione di oltre il 90% degli incidenti stradali che si verificano annualmente. Incidenti provocati, nella maggior parte dei casi, da errori umani dovuti alla distrazione, a un elevato tasso alcolico presente nel sangue, o a uno stato emotivo alterato dell'automobilista; eventualità queste che non potranno in alcun modo toccare le auto automatizzate.

La problematica che si vuole trattare in tale sede tuttavia riguarda il restante 10% dei casi di sinistri stradali che l'A.I. non riuscirà ad eliminare. È in tali situazioni che sorge il problema della responsabilità e in particolare quella penale, la quale si ricorda, ai sensi dell'articolo 27 Cost. deve essere personale.

La questione della responsabilità riveste una grande importanza, e di ciò se ne fa carico lo stesso Parlamento Europeo, il quale, nella risoluzione del 16 febbraio 2017, scrive che *“Grazie agli strabilianti progressi tecnologici dell'ultimo decennio, non solo oggi i robot sono in grado di svolgere attività tipicamente ed esclusivamente umane, ma lo sviluppo di determinate caratteristiche autonome e cognitive, ad esempio la capacità di apprendere dall'esperienza e di prendere decisioni quasi indipendenti, li ha resi sempre più simili ad agenti che interagiscono con l'ambiente circostante e sono in grado di alterarlo in modo significativo. Pertanto, la questione giuridica derivante dall'azione nociva di un robot diventa essenziale”*<sup>115</sup>.

Le problematiche nascono in realtà dalla circostanza per cui *“più i robot sono autonomi, meno possono essere considerati come meri strumenti nelle mani di altri attori (quali il fabbricante, l'operatore, il proprietario, l'utilizzatore, ecc.)”*<sup>116</sup>. Tale circostanza ingenera preoccupazioni, in quanto, chi li ha progettati potrebbe non essere sempre in grado di prevedere le reazioni che il robot stesso potrebbe sviluppare.

---

<sup>115</sup>Parlamento Europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit., pt. Z. La Risoluzione invero focalizza la sua attenzione esclusivamente alle ipotesi di responsabilità civile per i danni provocati dagli agenti intelligenti, tuttavia costituisce una valida base di ragionamento anche per le questioni di carattere penale.

<sup>116</sup> Parlamento Europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit, punto AB.

Ed è proprio in questo che consiste il cd. “*Responsability gap*” individuato dalla Risoluzione quale problema centrale della “*robot liability*”<sup>117</sup>.

Detto in altre parole, il nodo centrale del problema della responsabilità per i danni provocati dai robot consiste nel fatto che, tali danni, potrebbero essere provocati da azioni non programmate da colui che lo ha progettato, realizzandosi così un vuoto di responsabilità. Ci si chiede se esista un soggetto al quale attribuire la responsabilità per tali azioni, e nel caso si dovesse dare risposta affermativa, chi dovrebbe essere?

Pertanto ciò su cui il Parlamento Europeo si interroga nella Risoluzione sopra citata riguarda la possibilità di adattare le vigenti regole generali in materia di responsabilità alle nuove ipotesi di *robot liability*, ovvero, al contrario, se vi sia la necessità di predisporre nuove regole e nuove categorie giuridiche, al fine di risolvere la questione relativa ai vuoti di responsabilità nel caso di azioni od omissioni riconducibili, non all’azione di un essere umano specificatamente individuabile, ma piuttosto all’azione autonoma del robot<sup>118</sup>.

Al fine di impostare nella maniera più corretta ed esaustiva possibile l’analisi della questione della relativa alla responsabilità si ritiene necessario approfondire in primo luogo le cause dalle quali questa può sorgere, che possono essere ricondotte a due grandi filoni: danni derivanti da malfunzionamento e danni provocati da scelte che la macchina compie.

## **1.2 Distinzione delle cause dei danni: malfunzionamento e scelte**

La distinzione riguardante l’origine dei danni si presenta dovuta e necessaria stante le particolari modalità di funzionamento dei nuovi prodotti dell’intelligenza artificiale.

Come sopra analizzato, la caratteristica principale dei prodotti intelligenti è rappresentata dal fatto che essi sono in grado di apprendere dall’esperienza svolta su grandi masse di dati in base ad algoritmi, che indurranno la macchina a prendere una decisione piuttosto che un’altra, ad agire in un modo piuttosto che un altro<sup>119</sup>. Sorge così la possibilità che, le azioni e le decisioni prese in totale autonomia, provochino dei danni, facendo sorgere la conseguente questione della responsabilità.

---

<sup>117</sup> U. Ruffolo, *Intelligenza artificiale e responsabilità*, cit., p. 80.

<sup>118</sup> *Ivi*, p. 82.

<sup>119</sup> Si veda il cap. 1, par. 2.2 di questo lavoro relativo alle *Modalità di funzionamento*.

I diversi tipi di danno possono quindi nascere da due situazioni differenti ma allo stesso tempo accomunate dal modo in cui tali prodotti funzionano.

DANNI PROVOCATI DA MALFUNZIONAMENTO, EVOLUZIONE NON PREVISTA DELL'ALGORITMO.

Da un lato, le scelte che la macchina sarà portata ad assumere potrebbero essere diverse da quelle previste dai programmatori. Si dovrà quindi tenere in considerazione l'eventualità (seppur minima) che l'algoritmo si possa evolvere in una direzione sbagliata, tramite ragionamenti e processi non del tutto noti e calcolabili in sede di programmazione, provocando di conseguenza eventi dannosi. Ed è a questa eventualità che devono essere ricondotti quei danni che, nel corso della trattazione, verranno richiamati come "*danni derivanti da malfunzionamento*".

Si è visto precedentemente come fosse sufficiente oscurare un singolo pixel di un'immagine per far sì che l'algoritmo di machine learning alla base del software di riconoscimento immagini smettesse di classificarla correttamente<sup>120</sup>. Si tratta questo di un caso del tutto innocuo, tuttavia le modalità tramite cui l'errore si è generato sono comuni a tutti i sistemi di intelligenza artificiale. Così, è possibile che un drone militare, anziché colpire l'obiettivo abbatta dei civili o, un'automobile driverless anziché dare la precedenza ai pedoni, che regolarmente attraversino le strisce pedonali, li investa. In questo caso tuttavia, l'errore provocherebbe un danno di rilevante gravità.

Compito del legislatore penale sarà dunque quello di sondare la normativa esistente in materia e stabilire se sia possibile adattare le attuali norme alle situazioni prese in considerazione.

DANNI DERIVANTI DA SCELTE COMPIUTE IN SEDE DI PROGRAMMAZIONE

D'altra parte, i danni che tali prodotti possono causare, possono sorgere da una causa differente. I *danni derivanti da scelte* sono provocati non da un malfunzionamento, inteso in questo senso come un'evoluzione atipica dell'algoritmo, ma piuttosto da una scelta originaria effettuata in sede di programmazione. Non appare atipica l'eventualità secondo la quale, i programmatori, debbano risolvere un difficile dilemma etico: in situazioni del tutto imprevedibili e inevitabili<sup>121</sup>, in cui qualsiasi scelta si possa prendere comporterà

---

<sup>120</sup>J. Su, D. Vasconcellos Varagas, K. Sakurai, *One Pixel attack for fooling deep neural networks*, cit., pp. 828 e ss.

<sup>121</sup> Si tratta delle situazioni riconducibili al c.d. *Dilemma del carrello*.

danni rilevanti, in quale direzione dovrà essere indirizzato l'algoritmo? Chi dovrà prendere effettivamente tale scelta?

Queste questioni si riflettono inevitabilmente sul piano della responsabilità, suscitando interrogativi di carattere non solamente giuridico, ma prima ancora etico. Si dovrà stabilire pertanto se sia possibile, in situazioni del genere, attribuire la responsabilità a qualcuno o se al contrario possa operare, anche in questo caso, una causa di esclusione della responsabilità.

## **2. DANNI DERIVANTI DA MALFUNZIONAMENTO**

Prima ancora che dei danni generati da un malfunzionamento, ossia da un'evoluzione atipica e non prevista dell'algoritmo di machine learning, si dovrà necessariamente prendere in considerazione, quale elemento imprescindibile per una disamina completa dell'argomento, l'effettivo rischio che tali danni possano essere provocati da una simile causa e le conseguenze relative alla sua percezione.

In questo paragrafo si vuole analizzare come il rischio rappresenti un fattore centrale da tenere in considerazione per una futura ed eventuale regolamentazione della materia. Si assiste oggi ad una intensificazione del rischio sia sul versante oggettivo sia su quello soggettivo: da un lato infatti, vi è una maggior presenza di occasioni di rischio connesse al modo di produrre e vivere, dall'altro lato vi è una più acuta percezione dei rischi stessi. Con la globalizzazione e lo sviluppo delle tecnologie le occasioni di danno non rivestono più carattere singolare, alla stregua di un incidente isolato che si è disposti a tollerare, ma al contrario, il danno nasce e si riproduce attraverso il prodotto, caratterizzandosi per la tragicità e l'assoluta imprevedibilità delle sue conseguenze<sup>122</sup>.

Pertanto, al fine di assicurare un elevato livello di tutela dei consumatori e degli utenti, in Italia è stata recepita con il Decreto legislativo, 06/09/2005 n° 206 la direttiva Europea n. 374/1985/CE riguardante la responsabilità da prodotti difettosi<sup>123</sup>. Si cercherà in tale sede di analizzare se la disciplina del Codice del consumo possa essere estesa anche ai danni provocati da prodotto robotico o se al contrario sia necessario adattarla alle mutate tendenze del mercato.

---

<sup>122</sup>C. Piergallini, *Danno da prodotto e responsabilità penale, profili dogmatici e politico-criminali*, Milano, Giuffrè, 2004, p. 38.

<sup>123</sup> *Codice del consumo*.

Tuttavia, in quest'ottica, appare necessaria una tutela anche penale stante la tipologia di danni che i nuovi prodotti risultano potenzialmente in grado di provocare.

## **2.1 La riduzione delle aspettative di infallibilità a seguito dell'incidente di Uber**

Gli incidenti possono essere provocati, in primo luogo, da danni derivanti dal malfunzionamento dei vari sistemi e strumenti tramite cui la macchina agisce.

Il pieno sviluppo dei sistemi di A.I. è infatti un percorso arduo e la tecnologia necessita di essere messa appunto prima che la stessa venga commercializzata al grande pubblico.

È stato proprio un difetto di funzionamento ad aver determinato, il 18 marzo 2018, il primo incidente mortale che ha visto coinvolta un'auto a guida autonoma.

Il caso, sicuramente emblematico, ha comportato una grande battuta di arresto per le aspettative riguardo tali prodotti.

L'incidente, nel caso di specie, ha portato al decesso di Elaine Herzeberg, la quale al momento dell'impatto stava attraversando la strada, al di fuori dell'attraversamento pedonale designato, spingendo una bicicletta carica di borse della spesa.

Due sono state in particolare le cause che non hanno permesso di rilevare l'impatto in tempo: da un lato la distrazione della conducente, la quale in quel momento stava guardando un video in streaming e non la strada, come previsto dal regolamento Uber<sup>124</sup>; dall'altro lato, ed è questo che qui interessa, un difetto di rilevamento dell'ostacolo e di notifica al pilota circa la necessità di effettuare la frenata di emergenza.

La Herzemberg infatti è stata rilevata dai sensori per la prima volta 6 secondi prima dell'impatto a 115 metri di distanza, ma per oltre 4 secondi il sistema di guida automatica non ha rilevato che fosse necessaria la frenata di emergenza<sup>125</sup>.

In ogni caso la causa determinante del tragico incidente è da rinvenire sicuramente in un problema di malfunzionamento, sia del software, sia dei sensori di rilevamento ostacoli con i quali l'auto era equipaggiata.

---

<sup>124</sup> National Transportation Safety Board Office of Public Affairs, *Preliminary Report Released for Crash Involving Pedestrian*, 24 maggio 2018.

<sup>125</sup> National Transportation Safety Board Office of Public Affairs, *Preliminary Report Released for Crash Involving Pedestrian*, 24/5/2018. Si parla di frenata di emergenza riservata al conducente in quanto la casa automobilistica Uber ha ritenuto di disabilitare la tecnologia computerizzata di frenata automatica al fine di ridurre il rischio potenziale di comportamento irregolare del veicolo.

Il sistema aveva infatti effettivamente rilevato la Herzeberg sei secondi prima dell'impatto, tuttavia la ha classificata prima come un oggetto sconosciuto, poi come un veicolo e infine come una bicicletta, e, secondo la logica computerizzata, ognuno di questi elementi avrebbe tenuto un diverso percorso e una diversa velocità. È a questa circostanza che, secondo il report effettuato dalla polizia di Temple, si deve il fatto che il sistema di guida automatica abbia ritenuto necessario notificare di effettuare la frenata di emergenza solo 1,3 secondi prima dell'impatto.

D'altra parte, l'auto era equipaggiata con sensori radar e lidar dotati di tecnologia particolarmente avanzata ma che, secondo l'avvocato Arrowood, specializzato in auto senza conducente in Arizona, non sarebbero stati del tutto idonei a fornire una copertura a 360 gradi attorno al veicolo. Uber aveva infatti montato un solo sensore Lidar sopra il tetto tale da impedire il rilevamento di ostacoli a livello del suolo creando così un punto cieco intorno al veicolo.

La vicenda, come già accennato sopra, ha provocato una forte battuta d'arresto per le aspettative riguardo le auto a guida autonoma<sup>126</sup>.

## **2.2 La società del rischio: introduzione al concetto di rischio**

L'incidente provocato dall'automobile driverless di Uber rappresenta sicuramente un caso emblematico. Lo scalpore provocato e l'attenzione mediatica ricevuta hanno condotto a una riduzione delle aspettative circa l'infallibilità di tali prodotti e un calo dei relativi investimenti, mettendo così in discussione gli ottimi risultati ottenuti in anni di sperimentazione.

Se per anni la collettività ha accettato e affrontato il rischio in nome del progresso, ad oggi è invece diventata sempre più consapevole dei pericoli derivanti dalla continuazione a oltranza dell'*industrialismo classico*<sup>127</sup>.

---

<sup>126</sup> Dopo la collisione mortale per la Herzeberg, Uber ha interrotto i test sperimentali nelle città di Temple, San Francisco, Pittsburgh e Toront (ripresi poi il 20 dicembre 2018 a seguito dell'autorizzazione del Dipartimento dei trasporti della Pennsylvania). Pochi giorni dopo, il 26 marzo, il Governatore Ducey ha inviato una lettera a Dara Khosrowshahi, CEO di Uber sospendendo i test delle auto nell'intero stato.

<sup>127</sup> C. Piergallini, *Danno da prodotto e responsabilità penale*, cit. p. 34. Per "Industrialismo classico" si fa riferimento alla società industriale, guidata dall'obiettivo di massimizzare il flusso di risorse materiali disponibili, sviluppando, a tale scopo, le tecnologie. Nell'industrialismo classico la collettività accetta i rischi e le conseguenze negative come prezzo necessario per la crescita economica e l'aumento del reddito.

Tuttavia, dire che si debba arrivare a una riduzione totale del rischio appare utopistico. La riduzione a zero del rischio infatti estrometterebbe dal sistema un'ampia gamma di attività produttive potenzialmente rischiose<sup>128</sup>.

Le attività pertanto, ad oggi, vengono autorizzate alla sola condizione che il calcolo delle probabilità che da esse derivino eventi rischiosi sia accettabile.

*“Così nella società moderna il calcolo delle probabilità diventa una delle componenti essenziali della normazione del rischio”*<sup>129</sup>.

Questo orientamento è stato recepito anche in Italia, ove, ai sensi dell'articolo 104 primo comma del Codice del Consumo<sup>130</sup>, un prodotto, per essere commercializzato, deve rivestire le caratteristiche di un prodotto sicuro.

Ossia un prodotto che, *“in condizioni di uso normali o ragionevolmente prevedibili, non presenti alcun rischio oppure presenti unicamente rischi minimi”*<sup>131</sup>.

Quando si parla di automobili a guida autonoma, e più in generale di prodotti dell'intelligenza artificiale, tuttavia la situazione appare più complicata.

Da un punto di vista statistico, si è ripetuto più volte nel corso della trattazione, le auto a guida autonoma si presentano più sicure delle auto tradizionali, si prevede infatti che esse porteranno una riduzione di oltre il 90% degli incidenti stradali.

Tuttavia, nel calcolo del rischio viene in rilievo, non esclusivamente la sua oggettiva consistenza (alla stregua di un fatto descritto dalla scienza), ma, prima ancora, la sua percezione da parte della collettività, data dai processi culturali e sociali della sua interpretazione.

*“La scienza diventa così sempre più necessaria ma nello stesso tempo sempre meno sufficiente per la definizione socialmente vincolante della verità”*<sup>132</sup>.

---

Per un approfondimento sul tema si rimanda a U. Beck, *La società del rischio, verso una seconda modernità*, Carocci Editore, Roma, 2000, pp. 14 e ss.

<sup>128</sup> C. Piergallini, *Danno da prodotto e responsabilità penale*, cit., pp. 34 e ss.

<sup>129</sup> C. Piergallini, *Danno da prodotto e responsabilità penale*, cit., p. 35.

<sup>130</sup> Dlgs n°206 del 06/09/2005 coordinato ed aggiornato, da ultimo, con le modifiche apportate dalla L. n.37 del 3 maggio 2019 e dalla L. n.31 del 12 aprile 2019.

<sup>131</sup> Codice del consumo. Si riporta qui l'articolo 103 c1 lett a) *“Ai fini del presente titolo si intende per prodotto sicuro: qualsiasi prodotto, come definito all'articolo 3, comma 1, lettera e), che, in condizioni di uso normali o ragionevolmente prevedibili, compresa la durata e, se del caso, la messa in servizio, l'installazione e la manutenzione, non presenti alcun rischio oppure presenti unicamente rischi minimi, compatibili con l'impiego del prodotto e considerati accettabili nell'osservanza di un livello elevato di tutela della salute e della sicurezza delle persone”*

<sup>132</sup>C. Piergallini, *Danno da prodotto e responsabilità penale*, cit. p. 17

Ci troviamo, citando il termine coniato dal sociologo Beck, nella *società del rischio*<sup>133</sup>. Ciò non significa che viviamo in un mondo più pericoloso di quello di prima, ma più semplicemente che il rischio ha ormai occupato una posizione centrale e dominante nella vita di ognuno di noi così come nel dibattito pubblico, perché oramai lo percepiamo ovunque. Ed è ovunque. *“La società del rischio è una società catastrofica. In essa lo stato di emergenza minaccia di diventare la norma”*<sup>134</sup>.

Per comprendere come si è giunti alla società del rischio appare utile ripercorrere l’analisi effettuata dal filosofo Hans Blumenberg della metafora lucreziana del “naufragio con spettatore”<sup>135</sup>.

Nell’esordio del secondo libro del *De rerum natura*<sup>136</sup>, Lucrezio proietta una potente metafora sulla relazione intercorrente tra rischio e sicurezza. Uno spettatore, ben saldo sulla terraferma, contempla il mare in burrasca osservando con distacco la tragedia di un naufragio che si consuma davanti ai suoi occhi. Al contrario di quanto si potrebbe pensare, egli non ne è turbato, ma piuttosto, è pervaso da una gioia intima che scaturisce nel confronto tra la sua sicurezza e la posizione di pericolo corso da chi si trova nel mezzo di quella tragedia.

La metafora è permeata dalla filosofia Epicurea, secondo la quale l’uomo saggio, al pari dello spettatore descritto da Lucrezio, *“è colui che riesce a porsi a distanza di sicurezza con un ostentato distacco rivolto a contemplare lo spettacolo quotidiano della vita e della morte che la natura giornalmente consuma”*<sup>137</sup>.

Come sopra accennato, ciò che qui interessa è che attraverso l’analisi della metafora lucreziana, Blumenberg traccia un excursus storico dell’evoluzione del concetto di rischio e del grado di accettazione dello stesso nella società nell’arco di due millenni di storia.

Il saggio parte dall’analisi delle società stanziali, chiuse e fortemente radicate nelle tradizioni. Queste temono fortemente il rischio, lo evitano e qualora ciò non sia possibile, ricorrono a schemi di comportamento determinati e già conosciuti. Società nelle quali, “i

---

<sup>133</sup> Espressione originale: *Risikogesellschaft*. Per un’analisi completa dell’argomento si veda U. Beck, *La società del rischio, verso una seconda modernità*, Carocci Editore, Roma, 2000.

<sup>134</sup> U. Beck, *op.cit.*, p. 31.

<sup>135</sup> H. Blumenberg, *Naufragio con spettatore. Paradigma di una metafora dell’esistenza*, Il Mulino, Bologna, 2001.

<sup>136</sup> Lucrezio, *De rerum natura*, trad. it. a cura di E. Cetrangolo, Sansoni, Firenze, 1969, II, 1-4.

<sup>137</sup> C. Piergallini, *Danno da prodotto e responsabilità penale*, cit., p. 2. “Distanza di Sicurezza” è il titolo del saggio di Bodei, che introduce l’opera di Blumenberg, *Naufragio con spettatore. Paradigma della metafora dell’esistenza*, cit.

*precetti dispensati dal gruppo costituiscono in definitiva la bussola in dotazione a ciascun membro*” per emarginare il rischio.<sup>138</sup> Sono queste, pertanto, le società che incarnano i valori dettati dalla filosofia epicurea e posseduti dallo *spettatore* del naufragio della metafora lucreziana.

La svolta nel rapporto con cui la società concepisce e affronta il rischio si manifesta a partire dalla *rivoluzione copernicana*. Il celebre motto di Pascal “*vous etes embarqué*”<sup>139</sup> tratteggia pienamente tale mutamento: richiama infatti una condizione umana che non ha la possibilità di ritirarsi dinnanzi al pericolo perché il rischio rientra ormai tra le componenti costitutive della società. Sul piano culturale e sociale tale mutamento indica il *progetto della modernità*.<sup>140</sup>

L’apice di tale mutato atteggiamento dinnanzi al rischio, viene raggiunto con l’illuminismo e le prime due rivoluzioni industriali. Nella fase di sviluppo della società industriale infatti, il rischio viene generalmente accettato sul piano sociale in quanto elemento inevitabile per raggiungere il progresso e quindi il benessere della società stessa. I rischi industriali esistono e sono riconosciuti ma essi risultano caratterizzati dalla riconoscibilità delle vittime e vengono attuati da misure sociali, fattori che hanno reso il costo del progresso socialmente più accettabile<sup>141</sup>.

Oggi, nella società post-moderna, tuttavia il *progetto della modernità* sembra destinato ad esaurirsi. Il rischio industriale è caratterizzato da una parte dall’ampiezza e dalla gravità delle conseguenze, dall’altra dall’indeterminatezza delle vittime, è in grado, potenzialmente, di colpire tutti gli uomini nella stessa maniera, senza distinzione, concetto che Beck definisce come “*democraticizzazione del rischio*”<sup>142</sup>. Nell’odierna società dunque gli eventi dannosi non consistono più in incidenti isolati che si è disposti a tollerare, ma risultano piuttosto caratterizzati da un surplus di risultato lesivo.

Ciò che preoccupa dunque, “*non è tanto il moltiplicarsi delle occasioni di danno, quanto il carattere non definito e incalcolabile delle sue dimensioni*”<sup>143</sup>.

---

<sup>138</sup>Così si esprime C. Piergallini, *Danno da prodotto e responsabilità penale*, cit., p. 3.

<sup>139</sup> B. Pascal, *Pensieri*, a cura di P. Serini, Einaudi, Torino, 1967. Il motto di Pascal viene ripreso da Blumenberg, *Naufragio*, cit., p. 39.

<sup>140</sup> Per un’analisi della tematica Cfr. P. Rossi Monti, *Paragone degli ingegneri moderni e post moderni*, Il Mulino, Bologna, 1989, pp. 39 e ss.

<sup>141</sup> C. Piergallini, *Danno da prodotto e responsabilità penale*, cit., pp. 5 e ss.

<sup>142</sup> Così U. Beck, *La società del rischio*, cit., p. 48.

<sup>143</sup> C. Piergallini, *Danno da prodotto e responsabilità penale*, cit., p. 39.

È in questo contesto che si colloca la *Società del rischio*.

Appare evidente come oggi, le opportunità di un'esistenza sicura e gratificante siano notevolmente maggiori rispetto al passato, ma al contempo è tuttavia aumentato il timore e la cautela verso il progresso. È un processo circolare, per fare un esempio ci si affida sempre di più a sistemi esperti ma allo stesso tempo, inspiegabilmente, aumenta la diffidenza sull'attendibilità degli stessi<sup>144</sup>.

Si tende così a parlare del "*Paradosso del bisogno di sicurezza*" in base al quale anche se le condizioni di vita non sono mai state statisticamente così sicure, l'eccesso di complessità nel mondo continua a disorientare<sup>145</sup>. In una società così fortemente organizzata e dinnanzi a prodotti così tecnologicamente complessi risulta difficile accertarsi delle cause da cui discende il pericolo, e ciò incrementa sempre di più un forte bisogno di sicurezza.

È per questo motivo che nella valutazione del rischio la percezione assume sempre più un ruolo determinante ed è in grado, indirettamente, di influire su una potenziale regolamentazione della materia.

Sotto questo punto di vista, l'incidente provocato dal veicolo autonomo in Arizona ha prodotto enormi conseguenze, facendo crollare la fiducia degli investitori e scoraggiando i produttori dal testare i propri veicoli su strade pubbliche.

Si tratta di un vero e proprio *bias* cognitivo, ossia un errore sistematico che il cervello tende a commettere in circostanze specifiche, indirizzandoci verso decisioni basate su preferenze intuitive piuttosto che sulle regole del giudizio razionale<sup>146</sup>.

Per comprendere meglio il funzionamento dei bias si pensi al fatto che una delle paure più avvertite dall'uomo è quella di volare. Nel 2016, il secondo anno più sicuro della storia dell'aviazione, quaranta milioni di voli di linea arrivò tranquillamente a destinazione. Solo 10 invece rimasero coinvolti in incidenti fatali, ossia lo 0,000025% della cifra complessiva. Le testate giornalistiche di tutto il mondo riportarono però solo questi, d'altronde i passeggeri giunti salvi a destinazione non fanno notizia.

---

<sup>144</sup> *Ivi*, p. 8.

<sup>145</sup> *Ibidem*.

<sup>146</sup> D. Kahneman, *Pensieri lenti e veloci*, Mondadori, Milano, 2012, pp. 8 e ss.

*“L’istinto della paura è talmente forte da spingere gli esseri umani di tutto il pianeta, a rimuovere dal proprio campo visivo i 40 milioni di aerei che ogni anno giungono a destinazione senza problemi”<sup>147</sup>.*

Lo stesso processo cognitivo si è innestato a seguito del noto incidente provocato dalla driverless car di Uber. Le automobili di diverse case produttrici hanno percorso in modalità autonoma oltre 566 mila km nel 2017 e nei diversi tragitti solo 63 sono stati gli interventi umani che si sono resi necessari, con una percentuale quindi di 0.01% di interventi<sup>148</sup>. L’incidente avvenuto nel primo trimestre del 2018, e la conseguente attenzione mediatica ricevuta, ha tuttavia offuscato tali dati mettendo al contrario in evidenza un unico fatto: l’auto a guida autonoma non è così sicura come invece annunciato più volte dalle case automobilistiche.

Il nostro rapporto con la tecnologia è infatti viziato da un paradosso: tendiamo a fidarci eccessivamente di ciò che non capiamo, ma non appena scopriamo che un algoritmo può sbagliare crolla drasticamente tutta la fiducia che avevamo riposto nel suo funzionamento. Si parla a tal proposito di *“Avversione per gli algoritmi: siamo più tolleranti con i nostri errori che con quelli di un algoritmo”<sup>149</sup>*. D’altronde non potrebbe essere diversamente, l’idea che delle macchine si sostituiscano agli uomini in compiti sempre più delicati che, almeno nella nostra iniziale percezione risultano essere indelegabili, quale il prendere decisioni, ci allarma: proviamo disagio nel vedere replicate funzioni che sentiamo legate alla nostra più intima essenza<sup>150</sup>.

Così nonostante si sia stimato che le automobili a guida autonoma siano potenzialmente in grado di ridurre il 90% degli incidenti stradali<sup>151</sup>, un’unica notizia relativa a un caso facente parte del restante 10% è stata in grado, da sola, di eliminare dalla memoria degli investitori tutti i chilometri percorsi in sicurezza.

Quanto detto mette in luce un aspetto fondamentale: la commercializzazione o meno di tali prodotti dipende a monte dalla percezione della gente. Più il rischio è avvertito meno

---

<sup>147</sup> H. Rosling, *Factfulness*, Rizzoli, Segrate, 2019, pp. 127 e ss.

<sup>148</sup> Questo quanto si deduce dall’analisi di dati diffusi annualmente dal regolatore dei trasporti della California (unici dati accessibili al pubblico)

<sup>149</sup> Così H. Fry, *Hello World. Essere umani al tempo del digitale*, cit., p. 30.

<sup>150</sup> A. Carcaterra, *Machinae autonome e decisione robotica*, cit., p. 33.

<sup>151</sup> World Health Organization, *Global status report on road safety*, 2018.

sono le probabilità che tale prodotto sia effettivamente commercializzato e di conseguenza regolamentato.

## 2.3 Tipologie di danno da prodotto

Si è sopra esposto come, nella società post-moderna, il rischio sia un elemento che riveste un ruolo ormai centrale e come l'evento dannoso spesso nasca e si riproduca attraverso il prodotto. È per questo motivo che quando si parla di danni derivanti dal malfunzionamento dei prodotti dell'intelligenza artificiale non si può prescindere dal richiamare, come punto di partenza, la normativa dettata per i danni derivanti da prodotto difettoso.

Prima di sondare la possibilità di estendere la relativa disciplina anche ai prodotti in grado di auto apprendere, risulta necessario analizzare i diversi tipi di danno da prodotto, ossia i diversi difetti che il prodotto può presentare. Lo scopo è quello di identificare se i danni provocati dai prodotti dell'intelligenza artificiale, prodotti che apprendono con l'esperienza, possano essere ricondotti a una delle presenti categorie, ovvero se, con lo sviluppo dell'A.I., sarà opportuno aggiungerne una nuova.

Sulla base della normativa<sup>152</sup> e dell'elaborazione giurisprudenziale<sup>153</sup> si possono individuare tre tipologie di difetto. Si tratta, in particolare, di una distinzione incentrata sull'origine del danno subito dai consumatori:

- a) il difetto di fabbricazione;
- b) il difetto di progettazione o di costruzione;
- c) il difetto di informazione.

### DIFETTO DI FABBRICAZIONE

---

<sup>152</sup> Codice del consumo, Dlgs n°206 del 06/09/2005, così come modificato dalla L. nr 37 del 3/5/2019 e dalla L. nr 31 del 12/4/2019. Si veda in particolare la Parte IV, *Sicurezza e qualità*, Art. 102-135.

<sup>153</sup> Così il caso Contergan conclusosi nel dicembre 1970 con una sentenza di archiviazione "condizionata" (a norma del §153a dello StPO) e riguardante la commercializzazione del talidomide, preparato farmaceutico assunto da numerose gestanti, che, nella maggior parte dei casi diedero luce a figli con malformazioni congenite. Questo fu il primo caso che segnò l'esordio del diritto penale sul terreno della responsabilità del produttore. Si cita anche, tra le sentenze pionieristiche della giurisprudenza italiana la sentenza Cass sez IV, 30 sett 1997 relativo alla morte di un uomo, il locatario di un immobile, per avvelenamento provocato dall'ossido di carbonio fuoriuscito da una caldaia affetta da difetti strutturali e di installazione. Si cita infine, la più recente sentenza della Cass, sezione III, del 26 giugno 2015, n. 13225.

L'articolo 117 comma 3 del codice del consumo dispone che *“Un prodotto e' difettoso se non offre la sicurezza offerta normalmente dagli altri esemplari della medesima serie”*<sup>154</sup>.

Quando si parla di difetto di fabbricazione si allude pertanto alla circostanza per cui i vizi sono presenti esclusivamente in uno o più prodotti di una serie, che per il resto si presenta perfettamente funzionale.

In altre parole, si tratta di una difformità del prodotto singolo rispetto agli altri esemplari della stessa serie. Tale difetto, nel caso in questione, è riconducibile normalmente a “*de-faillances*” dei macchinari produttivi o ad errori umani degli operatori.

Interessa qui sottolineare come in questo caso si parli di un rischio prevedibile, la cui frequenza è, di regola, statisticamente calcolabile<sup>155</sup>.

#### DIFETTO DI PROGETTAZIONE O COSTRUZIONE

Tali vizi, contrariamente ai precedenti, ineriscono all'intera serie prodotta: tutti i prodotti della stessa serie presentano lo stesso difetto.

Essi possono discendere da molteplici cause quali un'errata progettazione, o scelta dei materiali e delle tecniche di produzione ovvero da un'insufficiente sperimentazione.

Questa categoria di vizi “è pertanto dotata di una *“rilevante potenzialità lesiva a causa dell'intrinseca capacità di proliferazione del danno che accompagna la vita del prodotto”*<sup>156</sup>. A questa tipologia di difetti è riconducibile, ad esempio, lo scandalo che ha colpito la grande casa americana *Bridgestone-Firestone*. La società fu accusata di aver messo in commercio nell'agosto del 2000 pneumatici difettosi montati su fuoristrada di lusso che, a causa di un *design defect* erano, non di rado, soggetti a vere e proprie esplosioni, provocando così oltre 1400 incidenti automobilistici con 88 decessi<sup>157</sup>.

In relazione ai difetti di costruzione, il primo comma lettera b) del medesimo articolo stabilisce che le legittime aspettative di sicurezza devono essere commisurate *“all'uso al quale il prodotto può essere ragionevolmente destinato e i comportamenti che in relazione ad esso si possono ragionevolmente prevedere”*<sup>158</sup>.

La norma impone al produttore, nella fase dell'ideazione, progettazione e costruzione del bene, di prefigurarsi (secondo canoni di ragionevole prevedibilità) gli usi e i comportamenti possibili del consumatore, onde adottare tutte quelle misure o quegli accorgimenti

---

<sup>154</sup> Codice del consumo, Parte IV, “Sicurezza e qualità” art. 117 c. 3.

<sup>155</sup> C. Piergallini, *Danno da prodotto e responsabilità penale*, cit., p. 47.

<sup>156</sup> *Ibidem*.

<sup>157</sup> *Ibidem*.

<sup>158</sup> Codice del consumo, Parte IV, “Sicurezza e qualità” art. 117. c1 lett. b.

che siano idonei a soddisfare le legittime aspettative di sicurezza, ovvero che consentano la tendenziale idoneità del bene ad essere utilizzato in condizioni di sicurezza.

Ciò implica che un uso del bene anomalo, irregolare e comunque difforme dalle istruzioni da parte del destinatario interrompe il nesso causale tra una asserita difettosità del bene e l'evento di danno, e quindi esclude la riferibilità di questo alla responsabilità del produttore<sup>159</sup>.

#### DIFETTO DI INFORMAZIONE

Tali difetti consistono nel mettere in circolazione un prodotto senza accompagnarlo con adeguate informazioni relative alle modalità del suo impiego.

Dispone a tal proposito il comma 1 alla lettera a) che *“Le legittime aspettative di sicurezza vengono assolte altresì attraverso il «modo in cui il prodotto è stato immesso in circolazione, la sua presentazione le sue caratteristiche palesi, le istruzioni e le avvertenze»*<sup>160</sup>

In questa ipotesi dunque il prodotto è costruito senza difetti strutturali ma risulta difettoso in quanto omette di informare, o informa in modo incompleto, il consumatore circa le modalità del suo impiego e gli eventuali pericoli connessi al suo utilizzo.

#### DIFETTI DA RISCHIO DI SVILUPPO

A queste categorie di difetti, generalmente riconosciute, alcuni autori<sup>161</sup> ritengono doveroso aggiungerne una quarta, costituita dai difetti collegati al *rischio di sviluppo*.

Si tratta di quei difetti che, nel momento in cui il prodotto è stato messo in circolazione, non potevano essere riconosciuti dato lo stato della tecnica e della scienza.

Tale categoria viene ricavata da una lettura a contrario dell'articolo 118 primo comma lettera e) ove si prevede una causa di esclusione della responsabilità del produttore nel caso in cui *“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”*<sup>162</sup>.

---

<sup>159</sup> C. Piergallini, *Danno da prodotto e responsabilità penale*, cit., p. 48.

<sup>160</sup> Codice del consumo, Parte IV, “Sicurezza e qualità” art 117. c1 lett. a.

<sup>161</sup> Tra cui C. Piergallini, in *Danno da prodotto e responsabilità penale, profili dogmatici e politico-criminali*. cit., p. 47.

<sup>162</sup> Codice del consumo, Parte IV, “Sicurezza e qualità” art 118 c1 lett e). Per completezza espositiva si riporta qui che l'articolo 118 include, tra le altre cause di esclusione della responsabilità: “ 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, ne' lo ha fabbricato o distribuito nell'esercizio della sua attività professionale; d) se il difetto e' dovuto alla conformita' del prodotto a una norma

I danni che scaturiscono da tali vizi sono pertanto non solo imprevedibili ed inevitabili, ma altresì caratterizzati da un'accentuata invasività in quanto, al pari dei vizi da costruzione, riguarderanno l'intera serie prodotta.

Si tratta a questo proposito di stabilire se, e in quale misura, i danni derivanti dall'apprendimento automatico caratterizzante i prodotti dell'intelligenza artificiale possano essere ricondotti a tale categoria di vizi.

## **2.4 Danno da machine learning: il problema del *Responsability Gap***

Spesso, le modalità con le quali un prodotto dell'A.I. risulta in grado di apprendere dall'esperienza non sono del tutto note ai tecnici della materia, e cosa più importante, non sono del tutto prevedibili i risultati a cui condurrà tale apprendimento.

Il fatto che tali danni potrebbero essere provocati da azioni non programmate da colui che ha progettato il prodotto costituisce infatti una tra le maggiori preoccupazioni manifestate in sede europea in tema di *Robot liability*<sup>163</sup>.

Da un punto di vista prettamente letterale, prendendo in considerazione l'articolo 118 primo comma lettera e), in tali casi il produttore potrebbe invero andare esente da responsabilità. Nel caso in cui un prodotto dell'A.I. provochi un danno commettendo un'azione non originariamente programmata, è ben possibile ammettere che “*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*”.

La questione invero risulta essere più complessa. Un'analisi impostata dal solo punto di vista letterale può risultare fuorviante. Sarà necessario pertanto prendere in considerazione altri elementi, quale la ratio della norma, le finalità e il momento storico in cui la stessa è stata emanata.

---

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>163</sup> Parlamento Europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit., voce “Responsabilità”.

La causa di esclusione della responsabilità da cui si fa discendere la categoria dei vizi derivanti dal *rischio di sviluppo* è stata inserita nel nostro ordinamento nel 2005<sup>164</sup> recependo una direttiva europea emanata nel 1985<sup>165</sup>, anno in cui l'intelligenza artificiale e il machine learning rappresentavano concetti poco più che fantascientifici.

Ci si chiede pertanto se, con la sempre maggiore diffusione dell'A.I, la causa di esclusione della responsabilità sia da riferire anche a quei prodotti caratterizzati dal fatto di auto apprendere. Ciò invero appare difficilmente ammissibile in quanto, se così fosse, si verrebbe a creare una facile scappatoia per tutti coloro che producono e immettono sul mercato tali prodotti, che in un futuro più o meno prossimo, saranno sempre più presenti nella vita di tutti i giorni.

Si realizzerebbe in tal maniera quel *responsability gap* particolarmente temuto in sede europea.

Tale conclusione appare condivisa anche dal Parlamento europeo, secondo cui tale ipotesi di responsabilità potrebbe non essere sufficiente a fronte dei danni potenzialmente scaturiti dalle nuove forme di AI che *“imparerebbero in modo autonomo, in base alle esperienze diversificate di ciascuno”*<sup>166</sup>.

Tali considerazioni rendono evidente come sia avvertita la necessità di rivisitare le attuali norme che regolano la materia, adattando quelle preesistenti ovvero creandone di nuove. Il tema verrà trattato nello specifico nel capitolo successivo<sup>167</sup>.

---

<sup>164</sup>La causa di esclusione della responsabilità oggi prevista dal Codice del consumo, Parte IV, “Sicurezza e qualità” art 118 c1 lett e) è stata infatti inserita tramite il Dlgs n°206 del 06/09/2005.

<sup>165</sup>Si tratta della direttiva 1985/374/CEE. I provvedimenti normativi in materia sono difatti emanati in attuazione di direttive europee: Il primo complesso normativo si ha con il d.P.R. 224/1988, di attuazione della direttiva europea 374/1985/CE, che si rivolge particolarmente alla responsabilità per danni da prodotti difettosi e il relativo regime di responsabilità, nonché il decreto legislativo correttivo n. 25/2001, di attuazione della direttiva europea 34/1999/CE. Tutto questo disposto è poi stato convogliato nel 2005 all'interno del Codice del Consumo dal d. lgs. 206/2005.

<sup>166</sup>Così si esprime il Parlamento Europeo nella Risoluzione *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit. pt. Z.

<sup>167</sup> Si veda il cap. 3 par. 3 della presente analisi.

## 2.5 *Ethic by design* come possibile soluzione

Ci troviamo pertanto in presenza di nuove entità, le cui decisioni e risultati non costituiscono l'effetto diretto di azioni umane, ma derivano piuttosto da una serie di processi che, dopo l'istruzione della macchina, hanno autonoma capacità decisionale. Tali risultati in alcuni casi non appaiono giustificabili a posteriori, in quanto non è possibile comprendere come il sistema sia giunto ad assumere una certa decisione.

Non sempre tuttavia è così. Di fronte a risultati ingiustificati ed ingiustificabili ve ne sono altri che, seppur lontanamente, è possibile intravedere e di conseguenza contenere.

Il tema, in particolare, è stato affronto nelle “Ethic Guidelines for Trustworthy AI”, linee guida per la realizzazione di sistemi di A.I. affidabile redatte da un apposito Gruppo di Alto livello di Esperti istituiti dall'Unione Europea<sup>168</sup>.

Per comprendere a pieno l'ottica in cui si colloca la questione appare opportuno analizzare i risultati ottenuti a seguito di una ricerca condotta presso il Georgia Institute of Technology<sup>169</sup>.

Lo studio ha posto in evidenza come, nelle automobili a guida autonoma, i sistemi di riconoscimento delle immagini rilevino i pedoni dalla pelle più scura con un tasso di accuratezza ridotto del 5 per cento.

Tale distorsione tuttavia non risulta dovuta a difetti di rilevamento dei sensori ma appare piuttosto causata da due elementi: troppo pochi esempi di pedoni dalla pelle scura e troppa poca enfasi sull'apprendimento da quegli esempi<sup>170</sup>.

Il dato appare particolarmente significativo in quanto può comportare la presenza di bias, distorsioni e conflitti di interessi annidati negli algoritmi e nei loro processi decisionali impliciti. Si introdurrebbero così negli algoritmi, inavvertitamente e involontariamente, discriminazioni e asimmetrie<sup>171</sup>.

Ciò, per fare un esempio, potrebbe condurre alla conseguenza che, in caso di pericolo imminente, il sistema dell'automobile driverless potrebbe essere indotto a sacrificare le persone dalla pelle più scura, in quanto dall'esperienza potrebbe aver imparato che esse abbiano un valore minore di quelle dalla pelle più chiara.

---

<sup>168</sup> The High-Level Expert Group on AI, *Ethic Guidelines for Trustworthy AI*, cit.

<sup>169</sup> B.Wilson, J.Hoffman, J.Morgenstern, *Predictive Inequity in Object Detection*, in «[arxiv.org](https://arxiv.org)», 21 febbraio 2019.

<sup>170</sup> K. Hao, *Self-driving cars may be more likely to hit you if you have dark skin*, in «MIT technology review», 2019, pp. 1 e ss.

<sup>171</sup> B.Wilson, J.Hoffman, J.Morgenstern, *Predictive Inequity in Object Detection*, cit. p. 3 e ss.

Qualsiasi distorsione nei dati utilizzati per addestrare gli algoritmi viene infatti amplificata nel processo decisionale della macchina<sup>172</sup>.

Come accennato prima pertanto a fronte di risultati imprevedibili a cui il sistema dell'intelligenza artificiale può giungere, ve ne sono altri che, con le dovute accortezze, possono essere in qualche modo intravisti. La distorsione individuata dal Georgia Institute of Technology rappresenta una di esse.

*L'ethic by design* può rappresentare un modo per superare tali bias evitando così che il software possa *evolversi* nella direzione sbagliata e riducendo di conseguenza le occasioni di danno.

#### COS'È L'ETHIC BY DESIGN

Il concetto di *ethic by design* viene introdotto nelle “Ethic Guidelines for Trustworthy AI”. Esse partono dal presupposto per cui, accrescendo la fiducia negli utenti, vi sarà una maggior diffusione dell'A.I. Si ritiene infatti che “*un approccio etico affidabile costituisca la via maestra per accrescere la competitività dell'Europa nel settore dell'intelligenza artificiale*”<sup>173</sup>.

Nello stesso documento, nel capitolo relativo al *Realising Trustworthy AI*, vengono fissati sette requisiti che devono essere soddisfatti affinché un sistema di intelligenza artificiale possa essere considerato affidabile<sup>174</sup>:

- 1-supervisione umana
- 2-robustezza tecnica e sicurezza
- 3-privacy e governance dei dati
- 4-trasparenza
- 5-diversità e non discriminazione
- 6-il benessere sociale e ambientale.
- 7-responsabilità

Affinché tali requisiti, costituenti principi e valori generalmente accettati nella comunità europea, vengano effettivamente implementati e non rimangano pertanto *lettera morta*, le Linee guida introducono una serie di metodi sia a livello tecnico che organizzativo.<sup>175</sup>

---

<sup>172</sup> *Ivi*.

<sup>173</sup> The High-Level Expert Group on AI, *Ethic Guidelines for Trustworthy AI*, cit., Introduction.

<sup>174</sup> *Ivi*, p. 14.

<sup>175</sup> *Ivi*, p. 20.

Tra essi è menzionata *L'ethic (and rule of law) by design* rappresentata dall'esigenza che il sistema A.I risponda, già in sede di progettazione ai sette valori sopra menzionati.<sup>176</sup>

Prendendo ad esempio in considerazione il requisito, indicato nel punto nr 5, della non discriminazione, è stato notato come gli insiemi di dati utilizzati nei sistemi A.I., sia in fase di addestramento che di funzionamento, potrebbero risentire dell'inclusione di determinati pregiudizi storici involontari che, se non adeguatamente corretti, potrebbero condurre a discriminazioni indirette e non intenzionali<sup>177</sup>.

Tali pregiudizi, identificabili e discriminatori, pertanto non solo dovrebbero essere rimossi già in fase di addestramento quando possibile, ma non dovrebbero nemmeno essere inseriti nei dati forniti originariamente ai sistemi di A.I. In questa maniera, il bias rilevato dal Georgia institute of technology nelle auto senza conducente dovrebbe essere corretto inserendo, fin dal principio, set di dati più inclusivi, nel caso in questione tanti esempi di persone con la pelle più scura quanto di quelle con la pelle più chiara. Si garantirebbe in questa maniera il rispetto del principio di diversità e non discriminazione enunciato nella prima parte del documento<sup>178</sup>.

L'etich by design potrebbe in definitiva garantire che un sistema A.I. sia, fin dalla fase della progettazione, etico.

Per citare le parole del Professor Floridi *“l'etica digitale deve informare le strategie per lo sviluppo e l'uso delle tecnologie digitali fin dall'inizio, quando cambiare il corso dell'azione è più facile e meno costoso, in termini di risorse e impatto. Deve sedere al tavolo delle procedure politiche e decisionali fin dal primo giorno”*<sup>179</sup>

Si ricorda in conclusione che, in quanto linee guida, i requisiti o i metodi ivi menzionati sono assolutamente privi di efficacia vincolante, anche se, allo stesso tempo, rivestono una fondamentale importanza nei confronti di coloro che operano nel settore dell'A.I. o si servono di esso.

Viene, a tal proposito, più volte sottolineato come *“nulla in questo documento deve creare diritti o imporre obblighi legali nei confronti di terzi”*<sup>180</sup> tuttavia *“gli stakeholder impegnati al raggiungimento di un'AI degna di fiducia possono volontariamente scegliere di*

---

<sup>176</sup>Ivi, p. 21.

<sup>177</sup> Ivi, p. 13.

<sup>178</sup> Ivi, p. 14.

<sup>179</sup> L.Floridi, *Soft Ethics e la governance del digitale*, in «Philosophy and Technology», 1, 2018, p. 5.

<sup>180</sup> The High-Level Expert Group on AI, *Ethic Guidelines for Trustworthy AI*, cit., p. 2.

utilizzare le presenti Linee guida come metodo per rendere operativo l'impegno assunto"<sup>181</sup> raggiungendo così l'obiettivo di presentare sul mercato un'intelligenza artificiale affidabile.

In questa direzione appare muoversi anche il colosso europeo Atos che si occupa di servizi IT, ove, l'*head of strategy portfolio* John Hall ha sottolineato l'importanza di un approccio etico by design. Hall ha dichiarato infatti come *“le aziende dovrebbero lavorare sull'innovazione con modalità che hanno molte più probabilità di essere accettati. Il mancato rispetto del limite di accettabilità porterà probabilmente al rifiuto delle innovazioni e, di conseguenza, a investimenti sprecati”*<sup>182</sup>. È importante pertanto che le aziende dimostrino di agire in modo affidabile e trasparente qualora vogliano proporre sul mercato tecnologie di cui gli utenti potrebbero non essere pienamente sicuri.

### **3. DANNI DERIVANTI DA SCELTE: LA QUESTIONE ETICA**

#### **3.1 Il dilemma del carrello: legame tra problematiche etiche e giuridiche**

Oltre che da errori di funzionamento del software, eventi nocivi possono essere causati anche da scelte che la macchina stessa sarà portata a prendere.

La situazione tipica che viene in rilievo in tale sede è riconducibile al cd. *Dilemma del carrello*<sup>183</sup>, posto per la prima volta nel 1967, ma ritornato recentemente in auge nelle discussioni riguardanti le vetture a guida autonoma<sup>184</sup>.

Si pensi alla situazione in cui una automobile a guida autonoma si trovi improvvisamente davanti tre persone che hanno iniziato ad attraversare la strada. In base alla situazione, il

---

<sup>181</sup> *Ivi*, p. 5.

<sup>182</sup> M. Dand, *The Future of AI is Ethics by Design, Diversity, and Inclusion*, in «Becoming Human: Artificial Intelligence magazine», 5, 2019, p.1.

<sup>183</sup> Le formulazioni originali dell'apologo del tram si attribuiscono a P. Foot, *The Problem of Abortion and the Doctrine of the Double Effect*, 1967, in «Oxford Review», 5, pp. 5-15, e a J. J.Thomson, *The Trolley Problem*, in «The Yale Law Journal», 5, 1985, pp. 1395-1415.

<sup>184</sup> Così T. Worstall, *When Should Your Driverless Car From Google Be Allowed To Kill You?*, in «Forbes», 18 giugno 2014.

sistema di controllo del veicolo stima di avere solo due possibilità: “*investire i pedoni che si trovano sulla sua traiettoria o sterzare bruscamente, portando il veicolo fuori strada e mettendo così a serio repentaglio la vita del suo unico passeggero e proprietario*”<sup>185</sup>.

Si richiede in questo caso che il veicolo, a seconda degli algoritmi posti alla base del sistema di machine learning, per il tramite del quale la macchina è in grado di agire, compia una scelta. Ogni tipo di scelta consapevole, e a maggior ragione quelle che devono essere assunte in situazioni come quella sopra prospettata, presuppongono un’etica<sup>186</sup>, la cui individuazione appare un importante punto di riferimento per affrontare il problema della responsabilità ed eventuali cause di esclusione della stessa.

La questione che viene in rilievo, e di cui si tratterà specificatamente nel capitolo successivo, riguarda il fatto che, in situazioni quali quelle prospettate dal trolley problem, la qualifica di prodotto autonomo modifica sostanzialmente i termini entro cui si dovrà inquadrare il problema della responsabilità. Nelle automobili tradizionali infatti, automobili controllate esclusivamente dall’azione umana, tale scelta verrà compiuta direttamente dall’automobilista in una situazione di pericolo estremo e soprattutto attuale. Ciò comporta che, in sede di giudizio penale, potrebbe essere applicata la scriminante dello *stato di necessità*<sup>187</sup>. Al contrario, in un’automobile a guida autonoma, la scelta che la stessa implementerà sarà una decisione standardizzata e pre-programmata, il che tenderebbe ad escludere la possibilità di applicare lo stato di necessità<sup>188</sup>, comportando così problematiche sul piano giuridico, oltre che sicuramente sul piano etico.

L’obbiettivo che tale analisi si propone di raggiungere è infatti quello di ragionare su temi etico-giuridici sollevati dalle nuove tecnologie. Si sottolinea tuttavia, che le questioni etiche e quelle giuridiche non siano separate tra loro ma, al contrario, vadano di pari passo.

---

<sup>185</sup> G. Tamburini, *Autonomia delle macchine e filosofia dell’intelligenza artificiale*, in «Rivista di filosofia», 2, 2017, p. 268.

<sup>186</sup> I. Rahwan, *The Social Dilemma Of Driverless Cars*, in TEDxCambridge, settembre 2016.

<sup>187</sup> La scriminante è prevista dall’art. 54 c.p. e risulta applicabile anche in caso di commissione di reati colposi. Per completezza si riporta di seguito il testo completo dell’articolo: “*Non è punibile chi ha commesso il fatto per esservi stato costretto dalla necessità di salvare sé od altri dal pericolo attuale di un danno grave alla persona, pericolo da lui non volontariamente causato, né altrimenti evitabile, sempre che il fatto sia proporzionato al pericolo. Questa disposizione non si applica a chi ha un particolare dovere giuridico di esporsi al pericolo. La disposizione della prima parte di questo articolo si applica anche se lo stato di necessità è determinato dall’altrui minaccia; ma, in tal caso, del fatto commesso dalla persona minacciata risponde chi l’ha costretta a commetterlo*”.

<sup>188</sup> In questo caso l’impossibilità di applicare l’articolo 54 c.p. risulta dovuta principalmente all’assenza dell’*attualità del pericolo*.

La riflessione su temi etici nasce difatti spontaneamente non appena ci si imbatte in un quesito morale di difficile soluzione, in questo caso quello riconducibile al *trolley problem*<sup>189</sup>. Giova premettere a tal proposito, per riprendere il pensiero espresso dal Professor Painsi, che riflettere sui problemi etici sollevati dall'utilizzo di un determinato prodotto dell'A.I non significa demonizzarlo, ma piuttosto, comprendere come aiutare le persone ad arricchire la propria consapevolezza e la propria capacità di governare il futuro e la tecnologia che avremo a disposizione, la quale offre grandi opportunità, col rischio che, se le stesse non verranno comprese verteranno a favore di altri<sup>190</sup>.

Detto in altri termini, il fatto che l'utilizzo tali prodotti ci costringa ad affrontare scelte eticamente difficili non dovrebbe frenarci dal cogliere le opportunità economiche e sociali offerte dallo sviluppo delle nuove tecnologie. Sarà necessario piuttosto ragionare su tali problemi etici al fine di arricchire la nostra consapevolezza e governare al meglio il futuro.

### **3.2 Risoluzione del problema tramite teorie deontologiche e consequenzialistiche**

Tornando al dilemma del carrello due sono in particolare le tesi che si contendono il campo offrendo soluzioni diametralmente opposte: Il consequenzialismo e il liberalismo di Kant. Ragionando secondo la logica consequenzialistica<sup>191</sup>, il software dovrebbe essere indirizzato a compiere la scelta che conduca, nel suo complesso, alle conseguenze migliori, ossia risparmiare il numero maggiore di vite.

Il Consequenzialismo<sup>192</sup> consiste infatti nella “*negazione dell’esistenza di assoluti morali mediante l’affermazione che gli atti umani non possiedono una qualità morale intrinseca,*

---

<sup>189</sup>M. Sandel, *Giustizia, il nostro bene comune*, Feltrinelli editore, Milano, Giugno 2019, p. 37.

<sup>190</sup> Così si esprime G. Painsi in *Digital World*, conferenza del 28 Febbraio 2020.

<sup>191</sup> Giova premettere fin da subito che l'utilitarismo è una forma di consequenzialismo, ma non tutti i consequenzialisti sono utilitaristi. Così G. Samek Lodovici, *Utilità del bene: Jeremy Bentham, l'utilitarismo e il consequenzialismo*, Vita e pensiero, Milano, 2004, p. 25.

<sup>192</sup> Termine introdotto da G.E Anscombe, *Modern Moral Philosophy*, in «Philosophy», 124, 1958, pp. 1 e ss.; ora in *Virtue Ethics*, a cura di R. Crisp- M. Slote, Oxford University Press, Oxford 1997, pp. 26-42.

*bensì ricevono una qualificazione mora diversa a seconda delle conseguenze prodotte*<sup>193</sup>. Secondo Bentham<sup>194</sup> infatti i diritti morali risultano essere privi di qualsiasi valore se non quello di strumento idoneo al perseguimento e alla promozione dell'utilità aggregata<sup>195</sup>.

Ciò che rileva dunque, al fine di considerare una determinata azione come etica o meno, non sarà l'intenzione originaria da cui è mosso l'autore dell'azione, ma piuttosto il risultato finale raggiunto a seguito dell'azione stessa.

Seguendo questa logica pertanto, sterzare bruscamente mettendo a repentaglio la vita dell'unico passeggero della self-driving car al fine di salvare i tre pedoni distratti non è sbagliato: *“finchè l'effetto è buono, non importa com'è la causa”*<sup>196</sup>.

Le considerazioni sopra svolte, riguardanti la scelta di impostare il software in modo tale da salvare il numero maggiore di vite, risultano valide solamente nel caso in cui si prenda in considerazione esclusivamente la singola situazione, il singolo dilemma del carrello che si prospetta in quel determinato momento. In realtà, una rigida applicazione del consequenzialismo, condurrebbe ad ampliare il punto di vista, prendendo in considerazione non il numero di vite risparmiate all'interno di una singola situazione ma il numero di vite salvate complessivamente. Detto in altri termini, se l'avvento delle auto a guida autonoma condurrà ad una drastica riduzione degli incidenti stradali sarà necessario incentivare l'acquisto di tali veicoli, e per farlo, bisognerà pertanto assicurare all'acquirente che la sua vita sarà in ogni caso salvaguardata.

In ogni caso, con le teorie consequenzialistiche si realizza uno *spostamento* del punto di vista da cui si parte per elaborare la filosofia morale: si realizza una sostituzione dal punto di vista della prima persona al punto di vista della terza persona. L'eticità di un'azione non passerà dunque attraverso la valutazione circa l'intenzione interiore del soggetto agente, ma piuttosto la valutazione, realizzata da un osservatore esterno, concernerà la conformità di un atto a una norma, un obbligo o un dovere, che nel caso in questione è rappresentato dalla massimizzazione dell'utilità<sup>197</sup>.

---

<sup>193</sup> G. Samek Lodovici, *Utilità del bene: Jeremy Bentham, l'utilitarismo e il consequenzialismo*, cit. pp. 20 e ss.

<sup>194</sup> Si cita Bentham in quanto la sua teoria utilitaristica rappresenta una forma di consequenzialismo. Cfr. J. Bentham, *Deontologia*, a cura di S. Cremaschi, La Nuova Italia, Firenze 2000.

<sup>195</sup> P. Ligutti, *I diritti come fini. La riflessione di Amartya Sen tra consequenzialismo e teorie deontologiche*, in «Filosofia politica», 3, 2004, p. 464.

<sup>196</sup> J. Bentham, *Deontologia*, cit. p. 37.

<sup>197</sup> G. Samek Lodovici, *Utilità del bene, Jeremy Bentham, l'utilitarismo e il consequenzialismo*, cit., p. 22. che per quanto riguarda la sostituzione del punto di vista fa riferimento a G. Abbà, *Felicità, vita buona e virtù. Saggio di filosofia morale*, Las, Roma, 1995.

Certo è che le teorie utilitaristiche e, più in generale, le teorie consequenzialistiche non sono andate esenti da critiche e obiezioni. Tra le principali contestazioni si rileva innanzitutto che la massimizzazione dell'utilità aggregata può condurre a ledere in modo significativo i diritti e le libertà fondamentali degli individui considerati singolarmente difatti, seguendo in modo coerente tale logica *“Gli individui contano, ma solo nel senso che si deve tener conto delle preferenze di ciascuno, accanto quelle degli altri”*<sup>198</sup>.

L'opposta soluzione al dilemma del carrello viene prospettata dalle teorie deontologiche, di cui si analizza in particolare il liberalismo di Kant. Seguendo tale impostazione bisognerebbe lasciare che, in tali situazioni, il software sia impostato in modo tale che l'autovettura prosegua il suo corso, indipendentemente dalle conseguenze che si potrebbero verificare, ossia indipendentemente dal fatto che ciò conduca al decesso di una, dieci o cento persone. Non si dovrà pertanto, in alcun modo, impostare il software in modo tale che sterzi bruscamente mettendo così a repentaglio la vita del passeggero, poiché in questo caso l'azione, ossia la programmazione, condurrebbe alla violazione del dovere di non uccidere. Le teorie deontologiche (deontologia dal greco δέον -οντος (deontos) e λογία (loghìa), è letteralmente lo "studio del dovere") *pongono particolare enfasi sul rapporto esistente tra dovere e moralità dell'azione*<sup>199</sup>. Nella valutazione circa l'eticità di un'azione pertanto dovrebbero rilevare esclusivamente le intenzioni originali che muovono l'azione stessa, non anche il risultato prodotto.

Kant, rifugge dalle teorie utilitaristiche non solamente come fondamento dell'etica personale, ma altresì come fondamento delle stesse leggi. Agire eticamente equivale ad agire spinti dal dovere, in nome della legge morale, ossia *“agire in modo da trattare sempre l'umanità, nella tua stessa persona o nella persona di chiunque altro, non come un semplice mezzo, ma sempre nello stesso tempo come un fine”*<sup>200</sup>. Agire liberamente ed eticamente per Kant significa dunque agire in armonia con l'imperativo categorico, cioè una *“massima di cui potresti nello stesso tempo desiderare che diventi una legge universale”*<sup>201</sup>.

---

<sup>198</sup> M. Sandel, *Giustizia, il nostro bene comune*, cit., p. 46. Per un approfondimento sulle ulteriori critiche mosse alle teorie consequenzialistiche si veda sempre M. Sandel, *Giustizia, il nostro bene comune*, cit., pp. 46 e ss.

<sup>199</sup> Voce *“Deontological ethics”* in Encyclopaedia Britannica.

<sup>200</sup> I. Kant, *Fondazione della metafisica dei costumi*, Editori Laterza, Bari, 2018, p. 91.

<sup>201</sup> *Ivi*, p. 89.

Programmare di uccidere, seppur il numero minore di persone, non può essere assurda a legge universale, nel risolvere il problema del carrello bisognerebbe pertanto lasciare l'auto al suo corso, non contravvenendo così all'imperativo categorico<sup>202</sup>.

Da semplice esperimento mentale, il *trolley problem* sta assumendo una sempre maggiore concretezza specie nel settore in esame. Con la sua semplicità ci obbliga a riconoscere un aspetto importante della questione delle auto a guida autonoma o, più in generale, dei prodotti dell'intelligenza artificiale: “*siamo di fronte a tecnologie che toccano questioni legate al potere, alle nostre aspettative, al controllo e alla delega della responsabilità*”<sup>203</sup>.

### **3.3 Società produttrici, Stato o conducenti: chi dovrebbe decidere del carrello? Introduzione alle conseguenze sul piano della responsabilità**

Tale scelta in ogni caso non viene presa dalla macchina in completa autonomia.

Come più volte ripetuto nel corso di questa trattazione, il prodotto dell'intelligenza artificiale non dispone di un'intelligenza propria nonostante esso sia in grado di auto apprendere grazie all'esperienza, ossia di compiere azioni che non sono mai state programmate tra quelle possibili<sup>204</sup>.

Questo implica che la risoluzione del dilemma etico, alla base di una scelta difficile quale quella posta dal *dilemma del carrello*, spetterà all'uomo, non alla macchina. *La questione etica infatti non riguarda l'intelligenza artificiale in quanto tale ma coloro che la producono*<sup>205</sup>.

Tre soggetti a tal proposito sono stati individuati quali possibili autori della scelta stessa: le diverse case produttrici del software, gli individui a bordo dell'auto e, infine, i diversi stati tramite una regolamentazione.

---

<sup>202</sup> Liberamente tratto da M. Sandel, *Giustizia, il nostro bene comune*, cit., pp. 138 e ss.

<sup>203</sup> H. Fry, *Hello World. Essere umani al tempo del digitale*, cit., p. 125.

<sup>204</sup> Si veda L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, cit., pp. 165 e ss.

<sup>205</sup> Così si esprime il Professore A. Pessina, docente di filosofia morale all'università Cattolica di Milano, in relazione a “*Rome Call for AI Ethics*”: l'intesa etica siglata in Vaticano da Pontificia Accademia per la Vita, Microsoft, IBM e Fao. Pessina propone una riflessione a margine del discorso che Papa Francesco ha rivolto a conclusione dei lavori dell'assemblea della Pontificia Accademia per la vita, del cui direttivo il docente della Cattolica è membro.

Nonostante per il momento i diversi progetti istituzionali suggeriscano che debba essere lo stato, o meglio tutti gli stati nel loro complesso, tramite una regolamentazione comune, a disciplinare l'impostazione di una simile scelta etica, e di conseguenza la politica di controllo dell'automobile a guida autonoma, si ritiene necessaria un'analisi delle conseguenze che deriverebbero nel caso in cui la stessa scelta venga affidata alle case produttrici del software o agli stessi automobilisti. Ognuna di queste impostazioni infatti comporterebbe conseguenze differenti e non trascurabili in materia di responsabilità.

#### LA DECISIONE DELLE CASE PRODUTTRICI

Qualora le case produttrici del software, o dell'automobile, potessero prendere in autonomia tale decisione, svincolate pertanto da una qualsiasi regolamentazione esistente in materia, verosimilmente, o almeno in un primo momento, sarebbero indirizzate a programmare le stesse in modo tale da massimizzare la sicurezza dei loro clienti. Ciò, di riflesso, comporterebbe che in situazioni riconducibili a quelle prospettate dal *trolley problem*, si incrementerebbe il rischio per i pedoni riducendo di conseguenza la sicurezza pubblica. La scelta di salvaguardare sempre e comunque l'integrità dei passeggeri dell'auto a discapito di eventuali passanti sarebbe giustificata da una semplice logica di mercato: gli automobilisti, come emerso più volte in numerosi sondaggi<sup>206</sup>, non sarebbero disposti ad acquistare un prodotto che sia pronto a sacrificarli in caso di situazioni di emergenza, e pertanto, tale scelta risulterebbe l'unica in grado di massimizzare i profitti.

Così nel 2016, al Salone dell'automobile di Parigi, Von Hugo, un portavoce della Mercedes Benz ha risposto al dilemma del carrello, prospettatogli da un giornalista in maniera molto chiara *“Salvare chi è a bordo del veicolo”*<sup>207</sup>.

Con tale dichiarazione però, la Mercedes ha scelto il lato sbagliato dell'opinione pubblica. Nei giorni successivi diversi articoli stigmatizzavano la scelta della casa automobilistica<sup>208</sup>. In uno di questi si leggeva *“Le loro automobili si comporteranno esattamente come lo stereotipo del guidatore di auto di lusso europee, libero di poter fare ciò che gli pare”*<sup>209</sup>.

---

<sup>206</sup> Tra i quali si ricorda J.F. Bonnefon, A. Shariff, I. Rahwan, *The social dilemma of autonomous vehicles*, in «Science», 352, 2016, pp. 1573-1576.

<sup>207</sup> M. Taylor, *Self-Driving Mercedes-Benzes Will Prioritize Occupant Safety over Pedestrians*, in «Car and Driver», 7 ottobre 2016.

<sup>208</sup> Per un commento sulla questione si veda H. Fry, *Hello World. Essere umani al tempo del digitale*, cit., pp. 123 e ss.

<sup>209</sup> J. Kottke, *Mercedes' Solution to the Trolley Problem*, [kotte.org](http://kotte.org), 24 ottobre 2016.

L'avversione nei confronti di quanto dichiarato da Hugo ha in realtà un fondamento più profondo e sicuramente più risalente nel tempo.

Se infatti le case produttrici potessero autonomamente impostare le scelte etiche richieste alla macchina, si cadrebbe in quella che l'economista inglese William Forster Lloyd descriveva come la "*Tragedia dei beni comuni*".

Lloyd, nel 1833 pubblicò un opuscolo, "*Two Lectures on the Checks to Population*"<sup>210</sup>, in cui, osservando la ricorrente devastazione dei pascoli comuni (ossia non di proprietà privata) in Inghilterra, descriveva una situazione immaginaria in cui tre agricoltori dividono una terra nella quale pascolano le proprie pecore. Fintanto che ogni agricoltore avesse posseduto, per fare un esempio, tre pecore, la terra sarebbe prosperata, le pecore avrebbero avuto cibo a sufficienza e gli agricoltori avrebbero ottenuto un buon formaggio: si sarebbe realizzata così la miglior soluzione per il bene comune.

Lo scenario muterebbe però nel momento in cui, uno dei tre agricoltori, perseguendo interessi esclusivamente personali, aggiungesse al suo pascolo una quarta pecora, ottenendo così, egoisticamente, un maggior vantaggio. In termini immediati nessuno ne subirebbe un danno, tuttavia, tale scelta ingenererebbe negli altri due agricoltori il desiderio di prendere la stessa decisione, ottenendo anch'essi un vantaggio maggiore. Questo tuttavia porterebbe alla conseguenza che la terra, in tal maniera, verrebbe abusata ed impoverita a discapito di tutti gli agricoltori. Il bene comune verrebbe abusato e distrutto.

Mettendo da parte l'esempio dei pascoli, ciò che Lloyd intendeva mettere in luce è rappresentato dal fatto che, nel caso in cui ognuno prendesse decisioni egoistiche, individualmente razionali, la risorsa comune potrebbe essere esaurita o addirittura distrutta, a scapito di tutti.

Trasferendo ora il discorso sul piano delle automobili a guida autonoma, il bene comune descritto da Lloyd è da rinvenire nella sicurezza pubblica, mentre gli agricoltori sono rappresentati dalle società produttrici del software. Queste, optando per la scelta arbitraria e individualmente razionale di preferire la sicurezza dei propri clienti a discapito di quella

---

<sup>210</sup> W. F. Lloyd, *Two Lectures On the Checks to Population*, Nabu Press, Firenze, 2014. L'opuscolo originale fu pubblicato nel 1833.

di eventuali passanti, rischierebbero di diminuire il bene comune rappresentato dall'incolumità collettiva cadendo così, in quella che viene descritta come *Tragedia dei beni comuni*<sup>211</sup>.

La situazione prospettata da Lloyd mette quindi in evidenza come una regolamentazione in materia sia necessaria, poiché in assenza di essa, la sicurezza pubblica, lasciata in balia delle singole decisioni individuali, verrebbe messa a serio rischio.

#### LA DECISIONE DELLO STATO

La regolamentazione pertanto sembra rappresentare la soluzione ideale al fine di preservare il bene comune, rappresentato nel contesto in questione, dalla sicurezza pubblica, che verrebbe altrimenti sacrificata in una prospettiva di guadagno individuale.

Tuttavia, una soluzione del genere, sebbene risulti lineare in via teorica, nella pratica si dimostra particolarmente complessa. Iyad Rahwan a tal proposito, partendo dall'assunto lockiano per cui le regole e le leggi sono un riflesso dei valori della società<sup>212</sup>, ritiene necessario, come punto di partenza, sottoporre il quesito ai cittadini prima di stilare un eventuale progetto di regolamentazione della materia<sup>213</sup>. Sennonchè, l'esito di un primo sondaggio da lui effettuato ha posto in evidenza come sul tema si registra quello che viene comunemente descritto come "*Dilemma sociale*", una situazione nella quale la razionalità individuale genera irrazionalità collettiva<sup>214</sup>. Dal questionario infatti è emerso come la maggioranza delle persone, posta di fronte al *dilemma del carrello*, abbia ritenuto preferibile la soluzione consequenzialistica, mostrando così (razionalmente) uno spiccato interesse verso la minimizzazione del danno collettivo, ammettendo poi però, a seguito di un'ulteriore domanda, che sarebbero disposti ad acquistare una vettura a guida autonoma solo qualora la politica di controllo della stessa abbia come obiettivo quello di proteggere e salvaguardare ad ogni costo la vita di chi vi è a bordo. È in situazioni del genere di irrazionalità collettiva, che si cadrebbe, qualora le stesse persone sottoposte a sondaggio avessero il potere di compiere in prima persona tale scelta, nella cd. *tragedia dei beni comuni*.

---

<sup>211</sup> L'espressione venne utilizzata da Lloyd in *Two Lectures On the Checks to Population*, cit. Il dilemma su esaminato per la prima volta da G. Hardin, *The Tragedy of the Commons*, in «Science», 3859, 1968, pp. 1243-1248.

<sup>212</sup> Il pensiero è espresso con particolare chiarezza in J. Locke, *Due trattati sul governo*, a cura di L. Pareyson, UTET, Torino, 2010.

<sup>213</sup> J. F. Bonnefon, A. Shariff, I. Rahwan, *The social dilemma of autonomous vehicles*, cit. p. 1573.

<sup>214</sup> *Ibidem*

Al fine di stilare una regolamentazione efficiente della materia, è dunque necessario che tale dilemma sociale venga in qualche modo risolto: è necessaria pertanto una società coesa.

Non soddisfatto dei risultati ottenuti a seguito di questo primo sondaggio condotto a titolo privato, Iyad Rahwan nel 2017, creò “*The moral machine experiment*”, una piattaforma sperimentale online progettata per esplorare i dilemmi morali affrontati dai veicoli autonomi che ha raccolto 40 milioni di decisioni in dieci lingue da oltre 2 milioni di persone in 233 paesi e territori diversi.

Tale esperimento consiste nel prospettare ai partecipanti diversi scenari riconducibili a quelli proposti dal trolley problem, nei quali, di volta in volta, il partecipante è chiamato a compiere una scelta: salvare il bambino o l’anziano, un gruppo di persone o una donna incinta che attraversa la strada nonostante vi sia il semaforo rosso, una folla o un criminale e così via<sup>215</sup>.

I risultati hanno mostrato come un ristretto numero di principi morali siano sostanzialmente condivisi da tutto il campione degli intervistati: come salvare il maggior numero di persone o privilegiare le vite umane a danno degli animali o degli oggetti<sup>216</sup>. Il che rappresenta sicuramente un risultato significativo che potrebbe andare a costituire uno dei mattoni fondamentali della regolamentazione, e dell’etica alla sua base, delle auto a guida automatica. Tale conclusione appare infatti essere in linea con le regole proposte dal report 2017 redatto dalla Commissione etica sulla guida automatica e richiesto dal ministero dei trasporti e delle infrastrutture digitali del governo tedesco (unica direttiva governativa oggi a disposizione)<sup>217</sup>.

Ma più in generale, lo studio ha dimostrato come i principi morali degli uomini non siano mai universali ma mutino a seconda di diversi fattori quali cultura, condizioni economiche ed area geografica.

---

<sup>215</sup> Si tratta di un esperimento cui è possibile partecipare accedendo al sito [moralmachine.mit.edu](http://moralmachine.mit.edu)

<sup>216</sup> Questo quanto analizzato da J. F. Bonnefon, A. Shariff, I. Rahwan, *The social dilemma of autonomous vehicles*, cit. pp. 1573 e ss.

<sup>217</sup> Federal Minister of Transport and Digital Infrastructure, *Automated and connected driving*, report Giugno 2017.

In conclusione, quindi, lo studio rivela che non esiste una morale universale riguardo a ciò che consideriamo giusto o sbagliato per quanto riguarda le decisioni che dovrebbero prendere le auto a guida automatica. Sebbene sia difficile stimare la frequenza con cui queste situazioni si presenteranno nella realtà, stabilire quali debbano essere i principi etici di una tecnologia con un'ampia prospettiva di diffusione è un tema che interessa non solo ingegneri ed esperti di etica, ma tutti i potenziali futuri consumatori. In altri termini, se le linee guida sull'etica dell'intelligenza artificiale venissero tracciate solo da esperti senza il coinvolgimento del grande pubblico, risulterebbero del tutto inutili<sup>218</sup>. *“Non è mai accaduto nella storia dell'umanità che si assegnasse a una macchina la responsabilità di decidere chi debba vivere o morire. Presto questo potrebbe accadere in uno degli ambiti più quotidiani della nostra vita: il trasporto privato. Lo studio rappresenta dunque anche un modo molto diretto per mettere la popolazione e i decisori politici a contatto con le problematiche che si celano dietro l'utilizzo su larga scala di tecnologie tanto innovative quanto delicate”*.<sup>219</sup>

Il fatto che la ricerca non abbia riscontrato l'esistenza di una morale universale condivisa rende evidente quanto sia complesso declinare prospettive etiche all'interno dei processi decisionali automatici di una driverless car.

Per questo motivo è stato di recente suggerito, al fine di ovviare il problema, di affidare la scelta etica direttamente ai futuri utenti.

#### LA DECISIONE DEL CONDUCENTE

La scelta etica potrebbe essere infine impostata dagli stessi automobilisti. Un gruppo di giuristi e ricercatori italiani dell'Università di Bologna<sup>220</sup> hanno suggerito, nel 2017, l'adozione di un Ethical Knob, una sorta di manopola etica che consenta alla persona a bordo di scegliere, nella prospettiva di un incidente, tra tre diverse opzioni di guida: altruistica, egoistica ed imparziale<sup>221</sup>.

---

<sup>218</sup> J. F. Bonnefon, A. Shariff, I. Rahwan, *The social dilemma of autonomous vehicles*, cit., pp. 1575 e ss.

<sup>219</sup> F. Suman, *Dilemmi morali per le auto a guida autonoma*, in «Il Bo Live Università di Padova», novembre 2018.

<sup>220</sup> Più specificatamente i ricercatori dell'European University Institute di Firenze, G. Contissa e F. Lagioia insieme a G. Sartor, docente all'Università di Bologna.

<sup>221</sup> G. Contissa, F. Lagioia, e G. Sartor, *The Ethical Knob: ethically-customisable automated vehicles and the law*, in «Artificial Intelligence and Law», 25, 2017, pp. 365 e ss.

Nella prima modalità, la modalità altruistica, l'importanza della vita delle altre persone supera l'importanza della vita dei passeggeri. Pertanto, l'AV dovrebbe sempre sacrificare i propri passeggeri per salvare altre persone, quali pedoni o passanti<sup>222</sup>.

Nella seconda modalità, imparziale, le vite dei passeggeri AV sono sullo stesso piano delle vite di altre persone. La decisione su chi dovrebbe essere salvato o sacrificato verrà presa per motivi utilitaristici, ossia l'opzione che minimizza ad esempio il numero di morti. In caso di perfetto equilibrio (in cui il numero di passeggeri è uguale a quello di terzi, potrebbe esserci una presunzione a favore di passeggeri o di terzi, o addirittura una scelta casuale tra i due<sup>223</sup>.

Nella terza modalità, infine, la modalità egoistica, l'importanza della vita dei passeggeri supererà sempre e comunque l'importanza della vita delle altre persone. Pertanto, l'AV dovrebbe sempre agire in modo da sacrificare i pedoni o i passanti anziché i propri passeggeri<sup>224</sup>.

La manopola consentirebbe inoltre all'utente di modificare l'impostazione in base alle circostanze, ad esempio quando il conducente abbia un bambino a bordo. Adottando la soluzione dell'Ethical Knob, al pari delle automobili tradizionali, resterebbe affidata all'individuo la possibilità di adottare una prospettiva personale su come affrontare il *dilemma del carrello*, con la sola differenza che, nel caso di vetture autonome, tale scelta verrebbe assunta in ogni caso in anticipo, in una situazione priva di *pericolo imminente*<sup>225</sup>, con conseguenze non trascurabili sul piano della responsabilità.

Tale proposta infatti, sebbene vada in controtendenza rispetto alla dottrina maggioritaria, che ritiene invece che l'impostazione etica debba essere implementata in maniera analoga in tutte le auto a guida autonoma, costituisce un importante spunto di riflessione sul piano della responsabilità.

Tuttavia, anche la soluzione prospettata dai giuristi dell'Università di Bologna rischierebbe di cadere nella sopra citata *tragedia dei beni comuni*.

---

<sup>222</sup> *Ivi.* p. 369

<sup>223</sup> *Ivi.* p. 370

<sup>224</sup> *Ibidem.*

<sup>225</sup> Si fa riferimento in particolare all'articolo 54 c.p. che prevede lo stato di necessità quale causa di esclusione della responsabilità.

### 3.4 Una quarta via

Per completezza di indagine si ritiene necessario analizzare una quarta via, per ovviare al dilemma del carrello.

Partendo dall'assunto prospettato dalla prima legge di Asimov sulla robotica, per cui "*un robot non può recar danno a un essere umano né può permettere che, a causa del suo mancato intervento, un essere umano riceva danno*"<sup>226</sup> parte della dottrina sostiene che la vera questione non concerna tanto la risoluzione in senso consequenzialistico o deontologico del dilemma del carrello, ma riguardi piuttosto il fatto di evitare alla radice situazioni nelle quali alla macchina venga richiesto di optare tra due alternative che risultino una peggiore dell'altra.

Tale prospettiva implica che, se necessario, non si dovrebbe mai arrivare a sviluppare le automobili autonome di livello 5, nelle quali l'uomo non è altro che un ospite. Il controllo ultimo sui prodotti dell'A.I dovrebbe pertanto restare sempre nelle mani dell'uomo, il quale deve poter essere messo nelle condizioni di spegnere in ogni momento la macchina, e non lasciare che essa prenda in totale autonomia decisioni così eticamente controverse.

L'impostazione proposta sembra rinvenirsi anche nelle linee guida adottate dalla Germania in cui è più volte sottolineato come "*l'uomo deve sempre poter avere il potere di sottrarre il comando alla macchina*"<sup>227</sup>, sostenendo indirettamente che non sia possibile risolvere il problema del carrello con soluzioni standardizzate o pre-programmate.

---

<sup>226</sup> Si riportano per completezza le Tre leggi di Asimov sulla robotica, formulate nel 1941, così come modificate a seguito dell'introduzione della Legge 0, enunciata nel romanzo *I Robot e l'Impero*, pubblicato nel 1985, la quale recita: "*Un robot non può recare danno all'umanità, né può permettere che, a causa del proprio mancato intervento, l'umanità riceva danno.*" 1-*Un robot non può recar danno a un essere umano né può permettere che, a causa del proprio mancato intervento, un essere umano riceva danno. Purché questo non contrasti con la Legge Zero.* 2-*Un robot deve obbedire agli ordini impartiti dagli esseri umani, purché tali ordini non contravvengano alla Legge Zero e alla Prima Legge.* 3-*Un robot deve proteggere la propria esistenza, purché questa autodifesa non contrasti con la Legge Zero, la Prima Legge e la Seconda Legge.* I. Asimov, *Io Robot*, Mondadori, Milano, 2018.

<sup>227</sup> Federal Minister of Transport and Digital Infrastructure, *Automated and connected driving*, report Giugno 2017, cit. p.9.

In questa prospettiva, le questioni relative alla responsabilità penale sembrano ritornare le stesse delle auto tradizionali ove la scelta etica rimane attribuita in ultima analisi all'automobilista e ove lo stato di necessità può essere preso in considerazione, a determinate condizioni, come una valida causa di esclusione della responsabilità.

## **Capitolo 3:**

### **Regolamentazione della responsabilità:**

#### **stato dell'arte e prospettive**

## **1. IL CONTESTO GIURIDICO E SOCIALE DI RIFERIMENTO**

In tale capitolo si analizzeranno nello specifico le problematiche sollevate con riferimento al tema dell'attribuzione della responsabilità penale per i danni provocati dai prodotti dell'intelligenza artificiale.

Al fine di cogliere ogni aspetto rilevante della tematica trattata, nel corso dell'analisi si accennerà spesso a un doveroso parallelismo concettuale tra le proposte elaborate in sede Europea con riferimento all'allocazione della responsabilità civile per i danni provocati dagli agenti artificiali e le nuove dottrine penalistiche sviluppatesi intorno a tali temi, comparandole e affiancandole continuamente agli insegnamenti dei grandi filosofi del passato. Il parallelismo con le proposte relative all'ambito civile risulta doveroso anche al fine di far emergere le cause della maggior complessità delle problematiche attinenti all'allocazione della responsabilità penale per i danni provocati da tali agenti e, dunque, al fine di individuare la strada ideale che il legislatore potrebbe seguire al fine di conciliare istanze, principi, diritti e doveri diversi e spesso in contrasto tra loro.

Si analizzerà il problema della responsabilità penale distinguendo le cause e le fonti da cui potrebbe scaturire il danno, così come si è già fatto nel precedente capitolo: da un lato, le ipotesi di danno derivante da malfunzionamento dell'algoritmo di machine learning e, dall'altro, quelle di danno derivante da scelte operate in sede di programmazione dall'*uomo dietro la macchina*.

Infine, per ogni ipotesi e proposta elaborata si ricercherà da un lato il fondamento etico e dall'altro le ripercussioni e le conseguenze sociali che potrebbero derivare.

## **1.1. Principi costituzionali in materia penale e difficoltà nel ripercorrere le strade individuate dal Parlamento Europeo**

Le poche proposte presenti nel panorama globale e, più specificatamente europeo, si propongono come obiettivo la regolamentazione della responsabilità a livello esclusivamente civile, proponendo in particolare la creazione di nuove categorie soggettive ovvero, al contrario, l'estensibilità delle preesistenti normative alla mutata realtà sociale<sup>228</sup>.

A livello di regolamentazione penale, al contrario, nulla è stato proposto attraverso canali ufficiali. Questo dato non deve sorprendere: la disciplina penalistica, sia a livello costituzionale che sovranazionale, è corredata da un insieme di principi e garanzie aventi lo scopo di definire i presupposti, il contenuto e i limiti della potestà punitiva. La punizione nei confronti di coloro che hanno violato il precetto penale rappresenta infatti un interesse, non della Pubblica Amministrazione in quanto tale, ma piuttosto dello Stato considerato nella sua unità, essendo indispensabile affinché la comunità possa esistere e progredire<sup>229</sup>.

Per raggiungere tale scopo, e quindi per evitare che tale potere possa essere esercitato in via discrezionale dalla pubblica amministrazione, risulta necessario contenerlo entro limiti fissati, nell'ordinamento italiano, direttamente dalla Costituzione.

Tali principi rappresentano un severo vincolo cui il legislatore deve necessariamente sottostare al fine di evitare di cadere in censure di costituzionalità e, specie quando vengono in rilievo materie quali quella qui esaminata, risultano idonei a spiegare tutta la loro forza ponendo, dottrina e organi legislativi, dinnanzi a serie problematiche.

Così se già le proposte elaborate in sede europea relative all'allocazione della responsabilità civile risultano idonee a suscitare non poche perplessità, le stesse proposte, proiettate sul piano penalistico, sono destinate a trovare il proprio limite nei principi generali che sottendono al diritto penale, ampliando così le già numerose problematiche.

Si consideri a tal proposito che, nel momento in cui ci si propone di normare le ipotesi di responsabilità penale per i danni provocati da prodotti dell'intelligenza artificiale, le ordinarie categorie del diritto nonché i principi e le garanzie costituzionali che reggono la

---

<sup>228</sup> Cfr. Parlamento Europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit. pt. nr. 51.

<sup>229</sup> F. Antolisei, *Manuale di Diritto Penale, Parte Generale*, XVI ed., Giuffrè, Milano, 2003, p. 63.

materia tendono a indebolirsi. Come sottolineato dal Parlamento Europeo infatti, ci troviamo oggi dinnanzi a entità non umane ma *in grado di svolgere attività che tradizionalmente erano tipicamente ed esclusivamente umane*, ossia in grado di *interagire con l'ambiente circostante e alterarlo in modo significativo*<sup>230</sup>. Entità quindi che si collocano a metà strada tra l'essere considerati meri strumenti e l'essere considerati veri e propri agenti.

#### ESTENSIONE DELLA DISCIPLINA DELLA RESPONSABILITÀ DA PRODOTTO DIFETTOSO

Il primo dei principi costituzionali in materia suscettibile di compressione è pertanto il principio della *personalità della responsabilità penale*, sancito dall'articolo 27 primo comma della Costituzione<sup>231</sup>. Il principio va oltre la mera esclusione della responsabilità per fatto altrui, ma implica un tipo di imputazione che si risolve in un giudizio di rimprovero. Si impone così la verifica *che il fatto appartenga personalisticamente e quindi spiritualmente al soggetto*<sup>232</sup>, in altri termini, bisognerà verificare la presenza dell'elemento psicologico, di colpa o dolo, in capo all'agente.

Il contenuto del principio costituzionale è ripreso dal codice penale nell'articolo 42 ove dopo aver stabilito che *“Nessuno può essere punito per un'azione od omissione preveduta dalla legge come reato, se non l'ha commessa con coscienza e volontà”*, prevede, al comma terzo, un temperamento nella parte in cui si stabilisce che *“La legge determina i casi nei quali l'evento è posto altrimenti a carico dell'agente, come conseguenza della sua azione od omissione”*<sup>233</sup>. Si parla a tal proposito delle c.d. ipotesi di responsabilità oggettiva: casi nei quali l'evento viene imputato al soggetto sulla sola base del rapporto di causalità, indipendentemente dal concorso del dolo o della colpa. L'agente viene in tali

---

<sup>230</sup> Parlamento Europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, pt. Z.

<sup>231</sup> Si riporta per completezza l'intero testo dell'Articolo 27 Cost. *“La responsabilità penale è personale. L'imputato non è considerato colpevole sino alla condanna definitiva. Le pene non possono consistere in trattamenti contrari al senso di umanità e devono tendere alla rieducazione del condannato. Non è ammessa la pena di morte.”*

<sup>232</sup> G.Bettiol, *Sul diritto penale dell'atteggiamento interiore*, in «Rivista italiana di diritto e procedura penale», 1971, pp. 3.

<sup>233</sup> Si riporta di seguito il contenuto dell'articolo 42 c.p. *“Nessuno può essere punito per un'azione od omissione preveduta dalla legge come reato, se non l'ha commessa con coscienza e volontà. Nessuno può essere punito per un fatto preveduto dalla legge come delitto, se non l'ha commesso con dolo, salvo i casi di delitto preterintenzionale o colposo espressamente previsti dalla legge. La legge determina i casi nei quali l'evento è posto altrimenti a carico dell'agente, come conseguenza della sua azione od omissione. Nelle contravvenzioni ciascuno risponde della propria azione od omissione cosciente e volontaria, sia essa dolosa o colposa”*.

casi chiamato a rispondere dei risultati della sua azione anche se nessun rimprovero, nemmeno di semplice leggerezza, gli possa essere mosso<sup>234</sup>. Sebbene la stessa Corte Costituzionale abbia stabilito che l'art 42 terzo comma non contrasti con il principio della personalità della pena<sup>235</sup>, tale ipotesi di responsabilità è stata spesso oggetto di vivaci critiche in quanto residuo di concezioni giuridiche ormai superate<sup>236</sup>. In ogni caso le ipotesi di responsabilità oggettiva previste nel nostro ordinamento sono determinate e tassative<sup>237</sup> e sia dottrina che una parte della giurisprudenza avvertono e sottolineano l'esigenza di limitarne il più possibile i casi.

Se pertanto in sede Comunitaria, con riferimento all'ambito civilistico, si è parlato di estendere la disciplina della responsabilità per danno da prodotto difettoso anche ai prodotti di *machine learning*<sup>238</sup>, disciplina che prevede ipotesi di responsabilità oggettiva in capo al produttore del bene; con riferimento all'ambito penale, l'ipotesi di una estensione della categoria penalistica della responsabilità per danno da prodotto si presenterà particolarmente controversa stante il vincolo dettato dal principio della *personalità della responsabilità penale* nonché l'esigenza di limitare il più possibile i casi di responsabilità oggettiva. Si consideri infatti che i prodotti dell'intelligenza artificiale sono entità in grado di apprendere dall'esperienza e, il particolare meccanismo di machine learning tramite cui sono in grado di agire, rende le loro azioni e il loro processo decisionale non del tutto prevedibile ab origine dai programmatori.

Diviene così infinitamente più complessa non solo la ricostruzione della trama causale del comportamento della macchina, ma anche l'individuazione del grado di diligenza richiesta al produttore. Si tratta in definitiva di aspetti che rendono difficoltoso attribuire la

---

<sup>234</sup> Così F. Antolisei, *Manuale di Diritto Penale, Parte Generale*, cit, pp. 388 e ss.

<sup>235</sup> Così la Corte Costituzionale, con la sentenza n. 364 del 24/3/1988. L'orientamento espresso si fonda sulla motivazione secondo la quale la responsabilità oggettiva pur prescindendo da un'indagine sull'atteggiamento psichico del soggetto, non prescinde invece dalla prova del nesso eziologico e della *suitas*. Non può essere quindi confusa con la responsabilità per fatto di terzi o per eventi al di fuori di qualsiasi potere di controllo dell'agente. Di conseguenza non contrasta di per sé con il principio della *Personalità della pena*.

<sup>236</sup> In questo senso si è espresso tra gli altri E. Mezger, *Diritto Penale (Strafrecht)*, trad. it. dell'Avv. Mandalari, CEDAM, Padova, 1935, p. 277.

<sup>237</sup> Si tratta del delitto preterintenzionale, delitti aggravati dall'evento e reati commessi a mezzo della stampa.

<sup>238</sup> Si fa sempre riferimento alla risoluzione del Parlamento Europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit., pt. AE.

responsabilità per i danni provocati da agenti intelligenti direttamente ai produttori o programmatori e portano a chiedersi se effettivamente non si stia parlando in questi casi di responsabilità per fatto altrui.

#### ATTRIBUZIONE DELLA RESPONSABILITÀ DIRETTAMENTE AL ROBOT

Non appare agevole tuttavia neppure percorrere l'opposta strada prospettata dalla Risoluzione del Febbraio 2017: attribuire la responsabilità direttamente al robot prevedendo a tal proposito la creazione di uno status giuridico ad hoc, rappresentato dalla categoria della c.d. *Personalità elettronica*<sup>239</sup>. A tal proposito va sottolineato come, non solo il diritto penale ha, da sempre, avuto come destinatari esclusivi le persone fisiche<sup>240</sup> ma anche come lo stesso articolo 27 comma terzo della Costituzione sancisca la c.d. *funzione rieducativa della pena*. Il fine perseguito dallo Stato tramite l'irrogazione sanzione penale è, tra gli altri, quello di recuperare il soggetto alla vita associata correggendone l'antisocialità offrendogli a tale scopo possibilità di recupero. Il principio Costituzionale formalizza quindi quell'esigenza già espressa nel Digesto dal giureconsulto Paolo secondo cui "*poena constitutor in emendationem hominum*"<sup>241</sup>. L'autore del fatto dimostra difatti la necessità di una guida, in modo tale da diventare elemento cooperante e integrato nella vita sociale. Nello stesso senso si muove anche la c.d. *teoria dell'espiazione* che attribuisce alla pena la funzione di purificazione dello spirito: la pena, secondo Platone, altro non rappresenta che *la medicina dell'anima*<sup>242</sup>.

Risulta difficile pertanto sostenere che, nel caso in cui il robot venga assoggettato a sanzione penale, la pena possa svolgere nei confronti dello stesso la funzione rieducativa la cui esigenza è affermata a livello Costituzionale.

---

<sup>239</sup> Parlamento Europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit., pt. 59, lett. f)

<sup>240</sup> Solo nel 2001, a fronte dell'esigenza di reprimere efficacemente gli abusi posti in essere direttamente dalle società e più in generale dalle persone giuridiche è stato emanato il Decreto legislativo 8 giugno 2001, n. 231, *Disciplina della responsabilità amministrativa delle persone giuridiche, delle società e delle associazioni anche prive di personalità giuridica*. Sebbene tale disciplina contenga nella stessa denominazione il termine "responsabilità amministrativa" impone delle previsioni sostanziali e procedurali di carattere sostanzialmente penale, la sua natura è infatti tuttora discussa. Per un approfondimento sul tema si veda con riferimento alla disciplina sostanziale G. Lattanzi, *Reati e Responsabilità degli Enti, guida al dlgs. 8 giugno 2001, n.231*, Giuffrè Editore, Milano, 2010 e con riferimento alla disciplina procedurale M. Ceresa-Castaldo, *Procedura Penale delle Società*, Giappichelli Editore, Torino, 2017.

<sup>241</sup> Trad. it. "*la pena è stabilita per il miglioramento degli uomini*".

<sup>242</sup> Platone, *Dialoghi Politici e lettere*, a cura di F. Adorno, UTET, Torino, 1988, 478 d.

Invero anche le altre funzioni della pena, richiamate dalla dottrina, quale la funzione retributiva ed intimidatoria, non appaiono attuabili nel momento in cui la stessa pena venga indirizzata a un soggetto diverso dalla persona fisica.

Si tratta invero delle medesime preoccupazioni che sono state sollevate con riferimento alla disciplina della responsabilità penale per le persone giuridiche<sup>243</sup> e che hanno contribuito al ritardo nell'emanazione di tale normativa. Si tratta tuttavia di ostacoli che sono stati superati, o aggirati, in materia di responsabilità degli enti, e pertanto si ritiene che non si possa escludere a priori che, anche nella tematica in esame, il legislatore possa intraprendere una simile strada.

Si consideri infine che, da una lettura più approfondita dell'articolo 25 della Costituzione, ribadito e precisato nel codice penale all'art. 1, viene ricavato il principio di *determinatezza e tassatività* del diritto penale, dal quale si deduce il c.d. *divieto di analogia in mala partem* del diritto penale. Il che implica che, nel caso in cui si opti per un'estensione delle norme attualmente esistenti nel nostro panorama normativo, sarà necessaria una modifica espressa da parte del legislatore.

## 1.2 La società delle Mangrovie

L'analisi dei principi costituzionali che vengono in rilievo in materia penale mette in luce una circostanza di rilevanza non secondaria: ci troviamo in quella che il filosofo Luciano Floridi descrive come *La società delle mangrovie*. Le mangrovie sono quelle particolari piante tropicali che crescono solamente nell'acqua salmastra, dovuta alla confluenza tra corsi fluviali e bacini oceanici, ossia la confluenza tra acque dolci e salate dando origine a qualcosa di nuovo e diverso<sup>244</sup>. Nel momento in cui si affronta il problema della regolamentazione della responsabilità penale per i danni provocati da prodotti dell'intelligenza artificiale, si scontrano interessi e principi fondamentali in contrasto tra loro, dando così luogo a quell'ambiente ibrido e salmastro in cui solo le mangrovie sono in grado di crescere e sopravvivere<sup>245</sup>. Come analizzato poco sopra infatti, individuare il soggetto

---

<sup>243</sup> Disciplina introdotta con il Decreto legislativo 8/6/2001, n. 231, *Disciplina della responsabilità amministrativa delle persone giuridiche, delle società e delle associazioni anche prive di personalità giuridica*.

<sup>244</sup> L. Floridi, *Soft Ethics e la governance del digitale*, cit. p.1.

<sup>245</sup> Così L. Floridi, *L'infosfera*, Onlife, cit. e L. Floridi, *Digitalisation, opportunities and challenges*, Chalmers University of Technology, Gothenburg, Sweden, March 2017.

responsabile nella figura del robot autore del danno attribuendogli la relativa responsabilità, così come tradizionalmente è previsto per le persone fisiche e giuridiche, verrebbe inevitabilmente a confliggere con l'impossibilità di attuare la funzione rieducativa nonché retributiva della pena nei confronti dei robot, ma allo stesso tempo, la punizione dei programmatori, che assicurerebbe l'individuazione di un soggetto responsabile, potrebbe venire a confliggere con il principio della personalità della responsabilità penale stante l'elevato grado di autonomia di tali agenti.

Una futura regolamentazione della materia dovrà pertanto incarnare quelle caratteristiche tipiche delle mangrovie, caratteristiche che le consentano di sopravvivere e operare efficacemente in un contesto ibrido nel quale si incontrano e si fondono diverse prerogative apparentemente, o realmente, in contrasto tra loro.

Se la sfida del *legislatore della società delle mangrovie* è quindi quella di trovare la strada per regolare la materia e scioglierne i relativi punti critici, sarà necessario che l'attenzione venga focalizzata sulla maniera migliore per ottimizzare i vantaggi e minimizzare i rischi connessi alle nuove tecnologie della società contemporanea.

I vantaggi da queste apportati sono stati esaminati nella prima parte della trattazione; così ad esempio, se la tecnologia driverless si dimostrasse anche solo lontanamente in grado di ridurre le vittime di incidenti stradali sarebbe immorale non cogliere tale opportunità<sup>246</sup>.

Sondando l'opposto versante, osserva Floridi, il principale rischio ad oggi è rappresentato dal sempre maggior potere degli algoritmi. Ciò comporta il fatto che troppo spesso siamo noi ad adattarci alle nuove tecnologie e non il contrario<sup>247</sup>.

Un ritardo nella regolamentazione e una conseguente impunità dei danni provocati da tali agenti potrebbe effettivamente concretare tale rischio. La tecnologia si muove troppo rapidamente rispetto alla legge, mentre invece è necessario che queste vadano di pari passo, è necessario ossia che l'intelligenza artificiale venga sviluppata nel pieno rispetto dei principi etici e costituzionali che informano la nostra società.

È pertanto evidente non solo la necessità di regolare tale materia, ma prima ancora che questa venga regolata secondo i valori rappresentativi della società e della comunità Europea in quanto, *non bisogna difendere ma promuovere; perché ci si difende contro qualcosa, ma si promuove ciò che si sa fare meglio*<sup>248</sup>. Detto in altri termini, non bisogna

---

<sup>246</sup> H. Fry, *Hello World. Essere Umani al tempo del digitale*, cit., p. 135.

<sup>247</sup> L. Floridi, *L'infosfera*, cit.

<sup>248</sup> L. Floridi, *Soft Ethics e la governance del digitale*, cit., pp. 1–8.

difendere i valori etici che informano il nostro diritto *da* un'asserita minaccia prospettata dalle nuove tecnologie, piuttosto bisogna promuoverli *con* l'avvento delle stesse.

In questo senso *l'Elogio ad Atene* di Tucidide dovrebbe incarnare la vera anima dell'Unione Europea nella parte in cui si afferma che “È giusto giudicare superiori per forza d'animo coloro che distinguono chiaramente le miserie e i piaceri, ma non per questo si lasciano spaventare dai pericoli”<sup>249</sup>.

L'Unione Europea infatti, in questo panorama riveste una posizione di primo piano: come ha già dimostrato nelle linee del GDPR<sup>250</sup>, attribuisce particolare rilevanza al ruolo dell'etica all'interno della regolamentazione, conciliando in tal maniera il diritto alla riservatezza con quello alla sicurezza e senza rifuggire a priori da un intervento normativo in materia stante il rischio di ledere possibili diritti. Lo stesso sarebbe auspicabile per quanto riguarda la normazione della responsabilità penale per i danni provocati dai prodotti dell'intelligenza artificiale, il rischio altrimenti, come sottolineato più volte nel corso della trattazione, potrebbe essere quello di diventare semplice consumatore di soluzioni sviluppate altrove: soluzioni che rischierebbero di non essere in linea con i principi garantiti a livello comunitario.

### 1.3 Necessità di distinguere le cause da cui scaturisce il danno

Se le proposte presentate a livello civile<sup>251</sup> si presentano come unitarie per tutti i tipi di danni provocati dagli agenti intelligenti, in ambito penale, al contrario, una distinzione relativa alle cause da cui sorgono gli eventi dannosi si presenta come necessaria al fine di inquadrare nella maniera più corretta la questione della responsabilità e tener conto di tutte le implicazioni da queste derivanti.

Come si è già accennato precedentemente<sup>252</sup> nel caso di danni derivanti da malfunzionamento si è dinnanzi a un'evoluzione anomala e inattesa del software di cui i programmatori potrebbero non esserne nemmeno lontanamente a conoscenza. In questi casi due sono

---

<sup>249</sup>Tucidide, *La guerra del Peloponneso*, trad. it. a cura di F. Ferrari, Milano, Rizzoli, 1985, Vol I, libro II, p. 328. Il brano citato è tratto dal discorso di Pericle in commemorazione dei caduti del primo anno di guerra (431 a.C.), ricostruito da Tucidide nel libro II della Guerra del Peloponneso.

<sup>250</sup> General Data Protection Regulation, Regolamento UE n. 2016/679

<sup>251</sup> Il riferimento va sempre alla Risoluzione Parlamento europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*. cit.

<sup>252</sup> Nel cap. 2 par. 1.2, *Necessità di una distinzione tra le cause dei danni*.

le strade prospettate dalla dottrina allo scopo di risolvere il problema relativo all'attribuzione della responsabilità penale: da un lato la creazione di un nuovo soggetto di diritto cui imputare direttamente l'azione, dall'altro l'attribuzione della stessa all'*uomo dietro la macchina*, prospettando in tal senso un'estensione della categoria penalistica della responsabilità per danno da prodotto.

Nel caso invece in cui l'evento dannoso sia stato provocato da una scelta del software, pianificata in precedenza dai programmatori o dai produttori e alla quale è sottesa una precisa decisione etica, la categoria della *soggettività elettronica*<sup>253</sup> risulta un'ipotesi non percorribile a priori. Al contrario in questi casi risulterà doveroso sondare il terreno delle cause di esclusione della responsabilità per verificare la possibilità di un'estensione delle stesse anche in tale ambito.

Si procederà pertanto analizzando le problematiche relative all'attribuzione della responsabilità tenendo fin da subito distinte le due ipotesi.

## **2. RESPONSABILITÀ PER I DANNI DERIVANTI DA MALFUNZIONAMENTO**

La Risoluzione emanata in sede Europea, pur non distinguendo le cause da cui il l'evento dannoso può scaturire, appare invero focalizzare l'attenzione principalmente sui danni provocati da un'evoluzione anomala dell'algoritmo di machine learning posto alla base del funzionamento dei sistemi di intelligenza artificiale.

### **2.1 Allocazione della responsabilità penale per danni causati da vetture semi-autonome e il control dilemma.**

Al fine di inquadrare correttamente la problematica della responsabilità penale per i danni derivanti da errori di malfunzionamento degli A.I. si ritiene necessario operare fin da subito una distinzione tra quei prodotti, automobili o droni militari, che si presentano come solo parzialmente autonomi e al contrario quelli completamente automatici. Per semplificare l'esposizione si è deciso di focalizzare l'attenzione, esclusivamente sulle

---

<sup>253</sup> Il termine è utilizzato nella Risoluzione Parlamento europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit., pt. 59 lett. f)

automobili a guida autonoma, ma, con i dovuti accorgimenti le stesse considerazioni potrebbero risultare valide anche con riferimento agli altri prodotti dell'intelligenza artificiale stante il medesimo meccanismo di funzionamento.

Come esposto nel primo capitolo di tale analisi, la circolazione dei veicoli autonomi è attualmente consentita dall'ordinamento italiano solamente a mezzi semiautonomi e con la presenza obbligatoria di un pilota umano nel veicolo in grado in qualsiasi momento di riprendere i comandi<sup>254</sup>, pertanto l'analisi che seguirà, relativa all'applicabilità del paradigma tradizionale del reato colposo d'evento agli eventi lesivi scaturenti dalla circolazione dei veicoli a guida completamente autonoma, sarà fondata su inquadramenti regolativi futuri ed ipotetici.

Prendendo dapprima in considerazione veicoli semi-autonomi, ossia mezzi ancora dotati di comandi manuali, si ritiene sempre possibile individuare un soggetto umano cui imputare gli eventi dannosi cagionati da malfunzionamenti: soggetto rappresentato dalla tradizionale figura del conducente<sup>255</sup>. Nelle auto semi-autonome, attualmente rappresentate da quelle vetture qualificate da un livello di automazione fino a 3<sup>256</sup>, è infatti necessaria una persona che monitori costantemente il funzionamento del veicolo, pronta a intervenire in caso di emergenza. Pertanto, finché permanga la materiale possibilità di riassumere in qualunque momento il controllo manuale del mezzo, si ritiene che tale categoria di soggetti, che da ora in poi verrà identificata col termine "*conducente potenziale*", assuma le stesse responsabilità del conducente tradizionale<sup>257</sup>. Entrambi infatti, governando il veicolo, una fonte di rischio altrimenti assente in natura, risultano sempre responsabili del controllo dello stesso, e saranno dunque investiti di una posizione di garanzia sia rispetto agli altri utenti della strada sia rispetto ai propri passeggeri<sup>258</sup>.

Sarebbe d'altronde arduo sostenere il contrario stante il fatto che ben difficilmente il legislatore potrebbe costruire intorno tali figure standard di cautela inferiori rispetto a quelli

---

<sup>254</sup> La circolazione è inoltre consentita solo per le sperimentazioni munite di apposita autorizzazione ministeriale

<sup>255</sup> Ovviamente anche rispetto ai mezzi semi-autonomi, così come già avviene con riferimento ai veicoli tradizionali (e come si analizzerà successivamente, con i dovuti accorgimenti anche nei confronti delle vetture full automation) è possibile configurare una responsabilità di singoli soggetti umani all'interno dell'apparato produttivo, secondo il paradigma penalistico del danno da prodotto.

<sup>256</sup> Dei 5 livelli di automazione si è discusso nel primo capitolo dell'analisi. Cfr. *SAE International's new standard J3016*, cit.

<sup>257</sup> Si evidenzia tuttavia come la nozione di "conducente" non sia attualmente definita dal Codice della strada.

<sup>258</sup> A. Cappellini, *Profili penalistici delle self-driving cars*, in «Diritto Penale Contemporaneo», 2, 2019, pp. 334 e ss.

suggeriti per gli automobilisti tradizionali creando di conseguenza un'area di non punibilità per i “conducenti potenziali” di tali mezzi<sup>259</sup>.

Il discorso si presenta in parte diverso rispetto alle vetture dotate di un livello di automazione pari a 4. Infatti, fintanto che il livello di automazione della *autonomous car* è basso, il conducente partecipa attivamente e continuamente all'attività del veicolo ben potendo essere considerato, in caso di imprudenza, responsabile per gli eventuali danni provocati; contrariamente, col crescere del livello di automazione, l'attività del *conducente potenziale* assumerà gradualmente le caratteristiche di una mera sorveglianza dell'attività autonoma del mezzo, senza la necessità di interventi attivi anche per lunghi periodi o tratti stradali<sup>260</sup>.

In definitiva, due risultano essere le strade potenzialmente percorribili: da un lato, proporre uno spazio di impunità per il conducente potenziale al pari di quanto, come si analizzerà successivamente, si potrebbe prospettare con riferimento alle automobili completamente automatiche di livello 5<sup>261</sup>, dall'altro lato, si potrebbe assimilare, con i dovuti accorgimenti, la responsabilità del *conducente potenziale* delle vetture parzialmente autonome di livello 4 a quella del conducente di vetture con un livello di automazione inferiore.

Per quanto riguarda quest'ultimo aspetto, ossia l'imputabilità al *conducente potenziale* di un evento lesivo causato da un malfunzionamento del veicolo con un livello di automazione pari a 4, è importante far luce sul fatto che, in tali casi, il paradigma della responsabilità, tradizionalmente di tipo ascrittivo commissivo, pare divenire omissivo. Il paradigma della responsabilità è tradizionalmente di tipo ascrittivo commissivo in quanto nella colpa stradale ordinaria, l'imprudenza, anche consistendo in una mancanza, ad esempio la mancanza della dovuta attenzione, non assume rilevanza di un'omissione in quanto tale ma piuttosto di un'omissione inserita nel quadro di una condotta più ampia: la guida del veicolo, avente indiscutibilmente e complessivamente carattere commissivo<sup>262</sup>. D'altronde, anche al di fuori dell'ambito della colpa stradale, la dottrina ha da

---

<sup>259</sup> *Ivi*.

<sup>260</sup> *Ivi*, p. 335.

<sup>261</sup> Si veda Cap. 3 par. 2.4 di questo lavoro.

<sup>262</sup> A. Cappellini, *Profili penalistici delle self-driving cars*, cit., p. 335.

sempre evidenziato in ogni tipo di colpa la presenza di un momento omissivo: ogni violazione cautelare può essere infatti vista come un'omissione di una cautela doverosa<sup>263</sup>.

Tale considerazione risulta valida non solo per le vetture tradizionali ma anche per quelle dotate dei più bassi livelli di automazione che si risolvono in mere funzioni di assistenza alla guida. Col crescere del supporto tecnico alla guida tuttavia, come precedentemente rilevato, l'attività del *conducente potenziale* si traduce in una mera attività di sorveglianza, pertanto, dal modello di evento lesivo cagionato commissivamente per colpa, si passerebbe a un modello imputativo omissivo per omessa sorveglianza, ossia per essere venuto meno all'obbligo di sorvegliare costantemente che il veicolo autonomo non incorra in errori o malfunzionamenti<sup>264</sup>.

In tale maniera dunque, resterebbe sostanzialmente inalterata quella tendenziale onnicomprensività della responsabilità del conducente per i fatti lesivi provocati nella circolazione stradale<sup>265</sup>: dal generale dovere di prudenza in capo al conducente, che oggi implica la responsabilità per colpa sostanzialmente per qualunque evento si verifichi, si passerebbe dunque ad un parimenti generale dovere di costante sorveglianza, rinvenendo nel conducente il capro espiatorio cui addossare la responsabilità di qualsivoglia danno prodotti<sup>266</sup>.

Né, d'altronde, è ragionevole attendersi che possa venire richiesto ai conducenti di tali vetture un grado di attenzione e sorveglianza alla strada minore, con conseguenti ricadute in termini di responsabilità, rispetto a quello richiesto per i conducenti di vetture tradizionali.

---

<sup>263</sup> Per tutti, cfr., F. Giunta, *Illiceità e colpevolezza nella responsabilità colposa*, CEDAM, Padova, 1993 pp. 90 ss. Sulla distinzione tra azione e omissione, cfr., P. Veneziani, *Regole cautelari "proprie" ed "improprie" nella prospettiva delle fattispecie colpose causalmente orientate*, CEDAM, Padova, 2003, pp. 44 ss.

<sup>264</sup> Diverso è il caso in cui la dovuta sorveglianza sia prestata, ma il conducente non si sia avveduto della necessità di un suo intervento, oppure abbia errato nel porlo in essere: in tal caso, probabilmente si dovrà parlare ancora di commissione, e non di omissione. cfr., A. Cappellini, *Profili penalistici delle self-driving cars*, cit., pp. 335-336.

<sup>265</sup> Responsabilità che attualmente continua a essere fortemente affermata a livello giurisprudenziale, che attualmente si mostra rigorosissima nella tutela del bene dell'incolumità degli utenti della strada, anche forzando i principi costituzionali quale quello di colpevolezza. Il riferimento teorico principale di tale tendenza rimane ancora infatti l'ormai risalente insegnamento di M. Duni, *Limiti all'obbligo di prevedere le imprudenze altrui*, in «Rivista Giuridica della Circolazione e dei Trasporti», 4, 1964, pp. 317-343. Secondo tale opinione, si ritiene che, per poter imputare un sinistro al conducente, sia sufficiente che l'evento sia anche solo astrattamente prevedibile: "l'obbligo di prevedere concerne tutto quello che non si vede e che può verificarsi da un momento all'altro: fatti naturali, fatti umani altrui, prudenti o imprudenti, fatti propri (come ad esempio, un malessere) ecc. Questa previsione costituisce l'elemento essenziale della prudenza". M. Duni, *op.cit.* p. 323.

<sup>266</sup> A. Cappellini, *Profili penalistici delle self-driving cars*, cit., p. 336

In un simile scenario, normativo e giurisprudenziale, appare pertanto irrealistico supporre l'introduzione, se non altro nel breve periodo, di una qualche limitazione legislativa della punibilità dei conducenti delle auto semi-autonome, anche se prossime alla completa automazione ossia di livello 4. Lo scetticismo ad abbandonare la figura del conducente quale «*parafulmine capace di attirare su di sé le colpe per qualunque evento occorso indipendentemente dalla sussistenza di un reale e ragionevolmente esigibile potere di controllo degli eventi*»<sup>267</sup> è in parte anche dovuta all'attuale contesto sociale e commerciale nel quale tali tecnologie sono inevitabilmente destinate ad inserirsi: si fa riferimento in proposito alla tendenziale ritrosia della società ad accettare il cambiamento di cui si è parlato in precedenza con riferimento alla *società del rischio*.

Una simile soluzione tuttavia, sebbene in linea con le attuali istanze legislative e sociali appena evidenziate, non pare tuttavia priva di punti di criticità sia nel breve che nel lungo periodo. La principale problematica sottostante tale indirizzo è stata definita in letteratura come *control dilemma*: nonostante le potenzialità tecnologiche offerte da un sempre crescente livello di autonomia della guida, il conducente non verrebbe mai liberato da responsabilità anche in caso di eventualità lesive imprevedibili e ingovernabili<sup>268</sup>. Ciò solleva un inevitabile contrasto non solo con il principio di colpevolezza, ma prima ancora con ragioni di carattere pratico: da un lato si assisterebbe alla perdita di tutti i vantaggi connessi a una circolazione stradale governata dalle nuove tecnologie<sup>269</sup> dall'altro lato, e soprattutto, una simile soluzione potrebbe addirittura generare rischi diversi e più numerosi rispetto quelli che l'operatività dell'obbligo di sorveglianza è effettivamente in grado di prevenire. Come hanno rilevato diversi studi scientifici infatti, l'attenzione, la capacità e la velocità di reazione umane appaiono essere fortemente limitate dallo svolgimento passivo di un compito di mera sorveglianza, contrariamente a quanto avviene nel momento in cui viene esercitato un ruolo attivo, quale la guida manuale<sup>270</sup>. In più si consideri

---

<sup>267</sup> A.Cappellini, *Profili penalistici delle self-driving cars*, cit., p. 337. con riferimento a E. Hilgendorf, *Automated Driving and the Law*, in *Robotics, Autonomics and the Law*, a cura di U.Seidel, Nomos, Baden-Baden, pp. 171-193.

<sup>268</sup> E. Hilgendorf, *Automated Driving and the Law*, cit., pp. 181 e ss.

<sup>269</sup> Si veda cap. 1, par. 3.2. di questo lavoro, relativo ai *vantaggi apportati dagli AV*.

<sup>270</sup> Tra i diversi studi che hanno condotto a tali risultati si cita: N. Schömig, V. Hargutt, A. Neukum, I. Petermann-Stock, I. Othersen, *The interaction between highly automated driving and the development of drowsiness*, in «*Procedia Manufacturing*», 3, 2015, p. 6657 e A.Hevelke e J.Nida-Rümelin, *Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis*, in «*Science and Engineering Ethics*», 21, 2015, p. 624.

che, proseguendo in questa direzione, si giungerebbe al paradosso secondo cui, conducenti sempre più disabituati a guidare manualmente, sarebbero invece chiamati a riprendere il controllo del veicolo nel momento più critico e difficile, ossia nelle situazioni di emergenza<sup>271</sup>. Va tenuta presente, infine, la problematica relativa ai “falsi positivi”, o “falsi pericoli”: situazioni nelle quali il *conducente potenziale*, erroneamente convinto di correre un rischio in realtà inesistente, intervenga manualmente in correzione della guida autonoma, ma provocando così, un sinistro che altrimenti non si sarebbe mai verificato<sup>272</sup>.

## 2.2 Allocazione della responsabilità per danni causati da vetture completamente autonome

Si è visto sopra come non poche siano le problematiche potenzialmente sollevabili con riferimento al problema della responsabilità qualora siano coinvolte automobili parzialmente automatiche di livello 4. Non deve pertanto sorprendere come il dibattito circa l’allocazione della responsabilità di prodotti intelligenti completamente autonomi presenti problemi ancora ulteriori e come, in tale materia la dottrina abbia presentato ipotesi e proposte di particolare audacia. La completa automazione, o per lo meno lo scopo che essa si prefigge di raggiungere, farebbe venir meno la figura del *conducente potenziale* umano. Il passeggero del mezzo, dato che di conducente non si può più parlare, potrà al più rispondere colposamente per mancata manutenzione, se ad essa sia riconducibile l’evento dannoso<sup>273</sup>.

La problematica vera e propria riguarda quegli errori di malfunzionamento dovuti a un’evoluzione anomala dell’algoritmo. Tali errori, seppur numericamente esigui, non potrebbero così essere imputati a un “conducente”, figura, come si è detto, assente in tale contesto. Chi risponderà dunque degli errori del robot?

---

<sup>271</sup> F. Douma, e S. Palodichuk, *Criminal Liability Issues Created by Autonomous Vehicles*, Santa Clara Law Review, 4, 2012, pp. 1157 e ss., che a tal proposito si spingono fino a parlare di un’intera generazione futura priva delle capacità di guida. F. Douma, e S. Palodichuk, *Criminal Liability Issues Created by Autonomous Vehicles*, cit., p. 1164.

<sup>272</sup> A. Cappellini, *Profili penalistici delle self-driving cars*, cit., p. 337.

<sup>273</sup> Ciò, ovviamente, risulta ammissibile solo qualora si parli di un passeggero che sia anche proprietario della vettura, o comunque in una posizione tale per cui su di lui gravi un qualche obbligo giuridico di assicurare la piena funzionalità del veicolo.

Si analizzeranno di seguito due opposte strade intraprese dalla dottrina: in primo luogo l'ipotesi, sicuramente più audace e fantasiosa, dell'attribuzione della responsabilità penale direttamente al prodotto dell'intelligenza artificiale; in secondo luogo si analizzerà la proposta di un'estensione della categoria penalistica del danno da prodotto anche agli eventi lesivi cagionati da tali agenti.

### **2.3 Una prima proposta: Robot come nuovo soggetto di diritto**

#### **LA PROPOSTA CHE CHIUDE IL CERCHIO DELLA QUARTA DELLA QUARTA RIVOLUZIONE**

Tra le diverse proposte citate anche in sede Europea, seppur con riferimento alla sola responsabilità civilistica, la creazione di una *personalità elettronica*, con la conseguente attribuzione della responsabilità direttamente in capo allo stesso robot, appare non solo la più sorprendente e fantasiosa ma soprattutto quella con una maggiore portata innovativa.

Si precisa fin da subito che la creazione di un nuovo soggetto di diritto per la quasi totalità degli studiosi della materia appare una strada difficilmente praticabile e poco auspicabile al fine di risolvere il problema della responsabilità<sup>274</sup>, tuttavia, la proposta apre numerosi spunti di riflessione non solo dal punto di vista prettamente giuridico ma prima ancora da una prospettiva puramente etica.

In effetti, attribuire alla macchina una propria soggettività porterebbe definitivamente a termine ciò che Floridi descrive come *la quarta rivoluzione*. Non si vuole far riferimento in tale sede alla quarta rivoluzione industriale di cui si è parlato nel primo capitolo, si tratta piuttosto di un diverso tipo di rivoluzione, quella che porta a una modifica della comprensione del mondo esterno e, di conseguenza, a una modifica della comprensione di noi stessi, di chi siamo<sup>275</sup>.

---

<sup>274</sup> Come si analizzerà nei paragrafi successivi, l'unico autore a sostegno della sua configurabilità, pur ragionando in termini di common law, è l'israeliano Gabriel Hallevy, di cui possiamo ricordare, oltre all'articolo dedicato espressamente al tema dei veicoli a guida autonoma: G.Hallevy, *Unmanned Vehicles: Subordination to Criminal Law under the Modern Concept of Criminal Liability*, in «Journal of Law, Information & Science», 2, 2011/2012, pp. 200-211, anche i più generali lavori: G.Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, Dordrecht, 2015; G.Hallevy *When Robots Kill. Artificial Intelligence under Criminal Law*, Northeastern University Press, Boston, 2013 e G.Hallevy *The Criminal Liability of Artificial Intelligence Entities, from Science Fiction to Legal Social Control*, in «Akron Intellectual Property Journal», 4, 2010, pp. 171-201. Il resto della dottrina, invece, nega ancora recisamente tale ipotesi.

<sup>275</sup> Così L. Floridi, *La quarta rivoluzione*, cit., p. 99.

La strada che ha condotto la società lungo le quattro rivoluzioni ha portato progressivamente verso una crisi della centralità dell'uomo e, considerare la macchina stessa nuovo soggetto di diritto e autrice del reato, sembrerebbe chiudere definitivamente tale cerchio.

Questo processo di decentralizzazione della figura umana ebbe inizio nel 1543, quando Copernico pubblicò il suo trattato sulla rotazione dei pianeti intorno al sole<sup>276</sup>. Se fino a quel momento si era portati a pensare che l'uomo, creatura prescelta da Dio, fosse al centro dell'Universo, la cosmologia eliocentrica prospettata da Copernico ci ha spinto a riconsiderare la posizione e il ruolo dell'essere umano all'interno di esso. Non eravamo più al centro del cosmo ma rimanevamo comunque al centro del regno biologico.

Anche questa certezza fu destinata a dissolversi nel momento in cui Darwin dimostrò, con *L'origine delle specie*<sup>277</sup>, che in realtà l'uomo non è sempre stato tale, ma altro non è che il frutto dell'evoluzione nel corso del tempo, da progenitori comuni ad altre specie, attraverso un processo di selezione naturale.

La specie umana, con la celebre affermazione di Cartesio "*cogito ergo sum*" rimaneva tuttavia padrone della propria mente, dei propri contenuti mentali, nonostante non fosse più al centro né dell'universo né del regno animale.

Fu Sigmund Freud a spazzare via anche quest'ultima illusione: i nostri pensieri, la nostra mente non è del tutto nostra in quanto tale, ma in quanto condizionata dal nostro passato. La mente è inconscia e soggetta a diversi meccanismi di difesa. *Siamo opachi a noi stessi*<sup>278</sup>.

Dopo queste tre rivoluzioni potevamo ancora ritenere che il nostro posto speciale nell'universo risiedesse nelle nostre superiori capacità di pensiero. L'intelligenza era una prerogativa esclusiva e gelosamente custodita dall'essere umano. Ma i germi della quarta rivoluzione furono gettati già nella metà del 1600 da Hobbes nel momento in cui nel *Leviatano* scrisse che "*La ragione non è che il calcolo (cioè l'addizione e la sottrazione) delle conseguenze al fine di contrassegnare e significare i nostri pensieri*"<sup>279</sup>. Per Hobbes quindi pensare equivaleva a ragionare, e ragionare altro non era che calcolare.

---

<sup>276</sup> N. Copernico, *Sulle rivoluzioni dei corpi celesti*, prima pubblicazione, 1543.

<sup>277</sup> C. Darwin, *L'origine della specie*, Rizzoli, Segrate, 2009.

<sup>278</sup> Così L. Floridi, *La quarta rivoluzione*, cit., p. 102.

<sup>279</sup> T. Hobbes, *Leviatano*, trad. it. a cura di G. Micheli, RCS libri, Milano, 2011, pp. 42 e ss.

Ciò che era ancora oscuro e non del tutto percepito divenne però evidente con Alan Turing, il vero padre della quarta rivoluzione. Con Turing, e il celebre test da lui predisposto<sup>280</sup>, l'essere umano ha definitivamente abbandonato l'idea di una posizione privilegiata anche nel regno del ragionamento logico, della capacità di processare informazioni e agire in modo intelligente<sup>281</sup>. Siamo stati costretti ad abbandonare una posizione che ritenevamo unicamente nostra. Si tratta invero, come emerso dai precedenti capitoli, di un passaggio in parte dettato dall'evoluzione dei tempi e sicuramente sofferto: pur avendo creato simili macchine e pur essendo affascinati dalla loro autonomia proviamo disagio nel veder replicate funzioni che sentiamo legate alla nostra più intima essenza “*dibattendoci, in questa contemplazione speculare, tra il mito di Narciso e l'eisoptrofobia, spaventati da uno specchio che ci rinvia alla nostra immagine riflessa*”<sup>282</sup>.

Individuare il robot come autore diretto dell'evento dannoso da lui causato, come accennato all'inizio del paragrafo, porterebbe definitivamente a conclusione tale processo: l'uomo non sarebbe più al centro nemmeno del diritto penale, diritto che da sempre ha avuto come esclusivo destinatario la persona fisica.

#### **DA COSA NASCE L'ESIGENZA DI RESPONSABILIZZARE DIRETTAMENTE IL ROBOT**

Prima di analizzare le implicazioni derivanti dall'attribuzione della responsabilità penale per i danni commessi dai prodotti dell'intelligenza artificiale direttamente agli stessi, appare opportuno interrogarsi sulle cause che hanno, a vario titolo, dato vita a tale esigenza. L'interrogativo circa l'eventualità di individuare un soggetto responsabile nuovo e diverso rispetto alla persona fisica nasce innanzitutto dall'attivazione di meccanismi, anche solo fattuali, di deresponsabilizzazione dell'agente umano<sup>283</sup>. L'utilizzo e il coinvolgimento dei sistemi di A.I. all'interno di talune attività innesca, inevitabilmente, un processo di *Alienazione della Responsabilità*, o per lo meno, di responsabilità morale.

Tale aspetto viene efficacemente descritto da Carla Bagnoli con riferimento all'impiego, nell'aeronautica, dei droni semiautonomi: “*Il comando dell'azione rimane all'essere*

---

<sup>280</sup> A. M. Turing, *Computing machinery and intelligence*, in «Mind», 59, 1950, pp. 433-460.

<sup>281</sup> Così L. Floridi, *La quarta rivoluzione*. cit., p. 103.

<sup>282</sup> A. Carcaterra, *Machinae autonomae e decisione robotica*, cit., p. 33.

<sup>283</sup> F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in «Diritto penale e Uomo», 5, 2019, p. 28.

*umano, ma questo non è materialmente l'agente dell'azione; anzi, occupa una posizione distante nel tempo e nello spazio e, talvolta, nemmeno ben identificabile.*

*Data la complessità della struttura gestionale e la distanza dell'azione, è difficile dire chi, alla fine, ha la responsabilità dell'azione. Soprattutto, poiché il comando è così frazionato, l'operatore non si identifica immediatamente con l'agente dell'azione, e ciò induce un tipo di alienazione con pericolosi effetti deresponsabilizzanti*<sup>284</sup>.

Tale effetto, di deresponsabilizzazione morale dell'agente umano rischierebbe di indebolire anche la tutela ai beni giuridici apportata e garantita dal diritto penale, e ciò, induce a ritenere che tale situazione potrebbe essere compensata dall'individuazione di un nuovo soggetto responsabile, soprattutto qualora questo nuovo soggetto sia *talmente autonomo da non poter essere considerato semplice strumento alle mani di altri attori*<sup>285</sup>.

Se oggi la proposta del Parlamento Europeo circa la creazione della c.d. personalità elettronica ci appare dirimpante rispetto alle regole base del diritto, si pensi che in realtà, nel dibattito giusfilosofico, la questione circa la possibilità di attribuire una responsabilità penale a entità diverse dall'uomo ha radici più radicate e profonde di quanto non saremmo portati a pensare. Queste furono gettate da Platone ne *Le Leggi*, ove attribuiva la responsabilità anche ad animali o cose<sup>286</sup> e implementate quando, nel XIII secolo Papa Innocenzo IV concesse lo status di personalità ai monasteri definendoli *sui iuris*. Nel 2001 poi, ciò che era rimasto per lungo tempo un mero dibattito assunse concretezza anche in Italia: si configurò un'ipotesi di responsabilità da reato in capo agli enti tramite l'emanazione del dlgs 231/2001.

L'ultima frontiera da abbattere riguarda pertanto l'equiparazione alle persone delle macchine di machine learning, o quantomeno di quelle maggiormente autonome, al fine dell'attribuzione in capo alle stesse della responsabilità penale.<sup>287</sup>

---

<sup>284</sup> C. Bagnoli, *Teoria della responsabilità*, Il Mulino, Bologna, 2019, pp. 78.

<sup>285</sup> Parlamento europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit., voce "Responsabilità" pt. AB.

<sup>286</sup> Platone, *Le leggi*, IX, 873e-874a, in, *Platone, Tutte le opere*, a cura di M. Cappuccilli, L'Erma di Bretschneider, Roma, 2003, p. 1335. Platone parla in particolare di un processo contro le cose "Se un oggetto inanimato toglie la vita a un uomo, almeno che non si tratti di un fulmine, proveniente da Dio, o perchè l'uomo urti su di esso, il parente più prossimo prenda come giudice il più vicino dei suoi vicini, e faccia espiazione per se e per tutta quanta la parentela; l'oggetto condannato sarà gettato fuori dei confini"

<sup>287</sup> Il dibattito in materia è stato inizialmente avviato dai filosofi del diritto e dai filosofi dell'informatica (cfr., tra gli altri, H. Jonas, *The Imperative of Responsibility. In search of an Ethics for the Technological*

Si ritiene in particolare che alla base dell'assioma del *machina delinquere non potest* risieda essenzialmente un pregiudizio di natura antropocentrica, lo stesso pregiudizio che si è per molto tempo opposto al riconoscimento di una responsabilità in capo alle persone giuridiche<sup>288</sup>.

D'altronde tale prospettiva non dovrebbe essere esclusa a priori in quanto, "una delle prerogative del razionalismo occidentale è la capacità di prendere le distanze dalle proprie tradizioni e allargare prospettive inizialmente circoscritte"<sup>289</sup>.

#### **OPINIONI FAVOREVOLI ALL'ATTRIBUZIONE DELLA RESPONSABILITÀ DIRETTAMENTE AL ROBOT: LA TESI DI HALLEVY**

Il principale sostenitore della tesi a sostegno dell'attribuzione al robot stesso della responsabilità penale per i danni provocati da uno sviluppo anomalo del software è l'israeliano Gabriel Hallevy, professore di diritto penale presso la facoltà di Giurisprudenza dell'Ono accademica College in Israele, che, nel suo celebre articolo *The Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control* conferma e analizza tale idea.

#### TRE MODELLI DI RESPONSABILITÀ

Hallevy individua in particolare tre modelli di responsabilità penale applicabile ai sistemi di intelligenza artificiale, il cui presupposto risiede nell'attribuzione agli stessi di una personalità giuridica. In tale maniera il professore israeliano si dichiara favorevole non solo al riconoscimento della personalità giuridica nei confronti di tali sistemi intelligenti,

---

Age, University of Chicago Press, Chicago, 1984; L.B. Solum, *Legal Personhood for Artificial Intelligences*, in «Illinois Public Law and Legal Theory Research Papers», 4, 2008, pp. 1231 e ss.; L. Floridi, J.W. Sanders, *On the Morality of Artificial Agents*, in «Minds and Machines», 3, 2004, pp. 349 ss.; B.C. Stahl, *Information, Ethics, and Computers: The Problem of Autonomous Moral Agents*, in «Minds and Machines», 14, 2004, pp. 67 ss.; B.C. Stahl, *Responsible Computers? A Case for Ascribing Quasi-Responsibility to Computers Independent of Personhood or Agency*, in «Ethics and Information Technology», 8, 2006, pp. 205 ss.; G. Sartor, *Gli agenti software: nuovi soggetti del ciberdiritto*, in «Contratto e impresa», 2, 2002, pp. 57 ss.) e si è di recente acceso anche tra gli studiosi della responsabilità civile (si veda, ad esempio, A. Santosuosso, C. Boscarato, F. Corleo, *Robot e diritto: una prima ricognizione*, in «Nuova giurisprudenza civile commentata», 2, 2012, pp. 497 ss.; D. Provolo, S. Riondato, F. Yenisey, *Genetics, Robotics, Law, Punishment*, Padova University Press, Padova, 2014, pp. 545 ss., nonché il volume a cura di U. Ruffolo, *Intelligenza artificiale e responsabilità*, Giuffrè editore, Milano, 2018.)

<sup>288</sup> Così ha sostenuto G.Hallevy, per la prima volta in *The Criminal Liability of Artificial Intelligence Entities, from Science Fiction to Legal Social Control*, cit., La tesi è stata tuttavia ribadita anche in scritti successivi tra cui si ricorda G.Hallevy, *Unmanned Vehicles: Subordination to Criminal Law under the Modern Concept of Criminal Liability*, cit., anche i più generali lavori: G.Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems*, cit., e G.Hallevy *When Robots Kill. Artificial Intelligence under Criminal Law*, cit. Il resto della dottrina, invece, nega ancora fortemente tale ipotesi.

<sup>289</sup> Così J.Habermas, *L'inclusione dell'altro, studi di teoria politica*, Feltrinelli editore, Milano, 2013, p. 222.

sulla stessa base del modello adottato da diversi ordinamenti giuridici per le società<sup>290</sup>, ma anche alla possibilità di una loro condanna penale obiettando così a quelle critiche facenti leva sul fatto che, un eventuale giudizio nei confronti del robot, non raggiungerebbe alcun risultato utile per la società o per la vittima: il robot non subirebbe né un effetto deterrente dalla condanna, né tanto meno sarebbe in grado di percepirne il disvalore<sup>291</sup>.

I tre modelli di responsabilità individuati nell'articolo si riferiscono a circostanze diverse, in particolare:

- Il primo di questi, *The perpetration through another*, è un modello che, non solo affonda le sue radici in un'epoca risalente nel tempo<sup>292</sup> ma che, ad oggi, viene generalmente riconosciuto anche nel diritto penale internazionale<sup>293</sup>. Afferma in particolare il principio secondo il quale, nel caso in cui venga commesso un fatto di reato attraverso l'azione o l'omissione di un diverso soggetto, il fatto viene attribuito al soggetto che utilizza l'autore materiale del crimine come un mero strumento, ferma restando la responsabilità dell'autore materiale nel caso in cui risulti penalmente perseguibile. Sul piano della responsabilità per i prodotti dell'intelligenza artificiale Hallevy considera quindi il robot privo di capacità cognitive e assimilabile all'incapace di intendere e volere, o al minore non imputabile: le sue capacità cognitive non sono sufficienti ad integrare i requisiti della mens rea. Il prodotto dell'A.I. si presenta dunque come il mero strumento di colui che architetta e commette il reato, il perpetrator. Come tale, è quest'ultimo, che Hallevy individua nella persona del programmatore o dell'utente, ad essere penalmente responsabile per l'actus reus dell'A.I.<sup>294</sup>. Questo modello tuttavia non risponde all'ipotesi che si vuole approfondire in tale sede. Esso fa esclusivo riferimento

---

<sup>290</sup> Si è parlato nei paragrafi precedenti delle difficoltà legate all'introduzione di un modello di responsabilità penale rivolto agli enti in corrispondenza dell'emanazione del Dlgs 231/2001.

<sup>291</sup> G. Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit. pp. 150 e ss.

<sup>292</sup>Le radici di tale modello affondano in particolare nel modello della responsabilità vicaria, riconosciuta dal diritto civile e penale e, un tempo, basata sul concetto di schiavitù e proprietà. Più in particolare, l'origine storica della responsabilità vicaria è inquadrabile nella tradizione giuridica romanistica, caratterizzandosi originariamente per la responsabilità del pater familias per i reati commessi dai membri appartenenti alla propria famiglia e dagli schiavi.

<sup>293</sup> Si fa riferimento in particolare all'art. 25 comma 1, lett. a) dello Statuto di Roma della Corte Penale Internazionale nella parte in cui si afferma la responsabilità anche di colui che ha commesso il fatto *through another person*.

<sup>294</sup>G. Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit. pp. 69 e ss., ma l'opinione è stata confermata anche negli altri scritti riconducibili al medesimo autore: G. Hallevy, *I, Robot—I, Criminal*—*When Science Fiction Becomes Reality: Legal Liability of AI Robots Committing Criminal Offenses*, cit.; G. Hallevy, *Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control*, cit.

a quelle ipotesi in cui il prodotto dell'intelligenza artificiale sia utilizzato come un mero strumento nelle mani di altri (penalmente responsabili ed imputabili) ma non anche alle situazioni in cui sia il prodotto stesso ad evolversi in una direzione non originariamente prevista in sede di programmazione

- Nemmeno il secondo modello, *The Natural Probable Consequence*, si sposa del tutto con la situazione specificamente presa in analisi. Sulla base di esso, i programmatori saranno individuati quali soggetti penalmente responsabili del reato commesso dall'A.I., nel caso in cui lo stesso si presenti come la conseguenza naturale o probabile di un loro comportamento negligente o doloso. Hallevy prende pertanto in considerazione la capacità dei programmatori di prevedere la potenziale commissione di reati da parte dei sistemi d'intelligenza artificiale, ed è tale capacità che funge da presupposto per l'attribuzione di una responsabilità penale<sup>295</sup>.

Per restare in linea con gli obbiettivi perseguiti dalla suddetta trattazione, ci si sofferma esclusivamente sull'ipotesi in cui l'azione criminosa posta in essere dalla macchina si presenti come la naturale conseguenza di un comportamento negligente del programmatore, comportamento quindi penalmente rilevante nella misura in cui, secondo un criterio di normale diligenza, il programmatore stesso avrebbe dovuto prevedere la possibilità che il reato fosse commesso<sup>296</sup>. Si noti come rispondono a tale scenario le ipotesi analizzate nel capitolo precedente, più in particolare, nel paragrafo relativo all'importanza di un approccio *Etic by design*. Come evidenziato dal Georgia Institute of Technology nel caso in cui, in fase di addestramento del software di machine learning, vengano inseriti set di dati non sufficientemente rappresentativi appare probabile che distorsioni o *bias* si annidino negli algoritmi e nei loro processi decisionali impliciti. Si tratta di un'ipotesi in grado di far sorgere una responsabilità colposa in capo al programmatore, in quanto, stante le particolari modalità di funzionamento dei sistemi di A.I., questo avrebbe ben potuto prevedere che tali distorsioni verranno poi amplificate nel processo decisionale della macchina e, ad esempio, in una situazione di pericolo imminente la stessa sarebbe portata a prendere una decisione non del tutto imparziale idonea a danneggiare una categoria precisa di persone. Affinché il programmatore possa pertanto andare esente da colpa sarà necessario correggere e superare tali *bias* attraverso

---

<sup>295</sup>G.Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit., pp. 75 e ss.

<sup>296</sup>*Ibidem*.

un approccio che si presenti etico fin dalla fase del design, inserendo un set di dati maggiormente inclusivo.

- Il terzo modello infine, *The Direct Liability*, appare corrispondere perfettamente alla problematica presa in considerazione in tale sede. Hallevy, nell'ultima parte del suo lavoro propone un modello di responsabilità che non presuppone alcuna dipendenza tra la condotta del programmatore, produttore o utilizzatore e quella posta in essere direttamente dal robot<sup>297</sup>. Ritiene, in altre parole, che il prodotto dell'intelligenza artificiale possa essere considerato direttamente responsabile per le azioni poste in essere dallo stesso, sempre che risultino integrati tutti gli elementi di cui il reato si compone.

L'autore, prendendo in considerazione la teoria della bipartizione del reato<sup>298</sup>, ritiene che lo stesso sia integrato qualora si verificano i due elementi dell'*actus reus*, elemento oggettivo, e della *mens rea*, elemento soggettivo. Si tratterà dunque di verificare se tali requisiti possano essere soddisfatti anche con riferimento ai sistemi di A.I. e, nel caso si riesca a dare risposta positiva, l'autore ritiene che non dovrebbero sussistere ragioni per non attribuire la responsabilità penale per l'azione commessa dalla macchina direttamente alla stessa.

#### ELEMENTO OGGETTIVO

Partendo dall'analisi dell'elemento oggettivo, Hallevy sostiene che, senza particolari problematiche, esso potrebbe essere integrato da un sistema di intelligenza artificiale. L'elemento oggettivo è di regola costituito da due componenti: la condotta e un evento tra le quali sussiste un legame, un *rapporto di causalità*, senza il quale l'avvenimento esterno non potrebbe considerarsi effetto della condotta<sup>299</sup>.

La condotta è rappresentata da un comportamento che si manifesta esteriormente tramite un'azione o un'omissione e, in entrambi i casi appare pacifico che questa possa essere posta in essere anche da un soggetto diverso dall'essere umano. Se il sistema di machine learning di un drone militare autonomo si evolve in una direzione non originariamente programmata e apre il fuoco sui civili anziché sull'obiettivo stabilito, il fatto di aver sparato ben può essere considerato un'azione dal punto di vista penalistico in quanto è da

---

<sup>297</sup>Ivi, pp. 82 e ss

<sup>298</sup>Cfr. F.Antolisei, *Manuale di Diritto Penale, Parte Generale*, cit, pp. 214 e ss.

<sup>299</sup>Ivi., pp. 219 e ss.

intendersi come un *movimento posto in essere dall'agente che si estrinseca nel mondo esteriore*<sup>300</sup>.

Allo stesso modo, nel caso in cui un'automobile a guida autonoma non si fermi in corrispondenza di un attraversamento pedonale e investa i passanti, la mancata frenata può essere considerata un'omissione in quanto, dal punto di vista del diritto penale il comportamento omissivo rileva nella misura in cui consista nel non compimento di azioni prescritte dall'ordine giuridico<sup>301</sup>.

Affinché l'elemento oggettivo possa considerarsi integrato è necessario che la condotta e l'evento lesivo siano collegati da un rapporto di causalità, senza il quale, il risultato esteriore non può essere considerato opera dell'agente e, di riflesso, non può essergli posto a carico<sup>302</sup>. Questa esigenza è sancita nel nostro ordinamento non solo a livello costituzionale, tramite il primo comma dell'articolo 27, ma altresì a livello codicistico: l'articolo 40 primo comma del codice penale stabilisce infatti che *“Nessuno può essere punito per un fatto preveduto dalla legge come reato, se l'evento dannoso o pericoloso, da cui dipende la esistenza del reato, non è conseguenza della sua azione od omissione.”* È a tale esigenza che si ricollega la teoria presa in analisi che attribuisce la responsabilità penale per le azioni poste in essere dai sistemi di intelligenza artificiale direttamente agli stessi; si rischierebbe altrimenti di violare i principi regolatori del diritto penale.

Tuttavia, nel tempo si sono sviluppate diverse teorie votate a stabilire cosa effettivamente occorra affinché un'azione possa essere considerata effettivamente causa di un determinato evento. Tra queste, oltre alla *teoria della causalità adeguata*, ha suscitato particolare interesse la teoria della *conditio sine qua non* secondo la quale deve considerarsi causa ogni singola condizione dell'evento, ossia ogni antecedente indispensabile senza il quale il risultato non si sarebbe verificato<sup>303</sup>. Seguendo alla lettera l'insegnamento di tale teoria non sembra sussistano ragioni affinché il programmatore o il produttore di un prodotto dell'intelligenza artificiale non possano essere ritenuti responsabili per il danno provocato

---

<sup>300</sup> *Ivi*, pp. 223-224

<sup>301</sup> G. Hallevy, *Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control*, cit., pp.187 e ss.; le stesse argomentazioni sono utilizzate da Hallevy in, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit., pp.39 e ss.

<sup>302</sup> F. Antolisei, *Manuale di Diritto Penale, Parte Generale*, cit., p. 219.

<sup>303</sup> La teoria è stata enunciata per la prima volta dal criminalista tedesco Von Buri, *Über Kausalität und deren Verantwortung*, Gebhardt, Lipsia, 1873, p. 33.

dal robot: il mancato sviluppo della tecnologia di machine learning, che permette alla macchina di apprendere dall'esperienza e assumere decisioni in autonomia, sarebbe stato sicuramente idoneo ad impedire la causazione del risultato lesivo. Sennonché alla stessa teoria della *conditio sine qua non* è stato mosso, e può tutt'ora muoversi, il rimprovero di apportare un'eccessiva dilatazione al del concetto di causalità, conducendo così a risultati contrastanti non solo con le esigenze del diritto ma prima ancora con lo stesso sentimento di giustizia<sup>304</sup>. Nonostante i diversi correttivi apportati nel tempo a tale teoria al fine di obbiettare alle critiche sopra esposte, la stessa appare in contrasto con lo stesso dettato codicistico, in particolare con l'articolo 41 del codice penale, il quale, nel secondo comma sembra effettivamente apporre un limite nella parte in cui afferma che "*Le cause sopravvenute escludono il rapporto di causalità quando sono state da sole sufficienti a determinare l'evento*"<sup>305</sup>. Se pertanto l'azione autonoma del robot si presenta da sola idonea a determinare l'evento, non essendo *ex ante* prevedibile, e quindi correggibile, in sede di programmazione, il nesso di causalità con la condotta del programmatore risulterà necessariamente interrotto. Si tratta in definitiva di un ulteriore elemento a favore della tesi che attribuisce la responsabilità penale direttamente al prodotto dell'intelligenza artificiale per i danni di lui provocati.

#### ELEMENTO SOGGETTIVO

La questione più delicata affrontata dal professor Hallevy, concerne tuttavia la possibilità di attribuire ad un A.I. l'elemento soggettivo del reato<sup>306</sup>. Secondo Hallevy, così come secondo le teorie generali del diritto, ai fini della configurabilità dell'elemento soggettivo o *della mens rea*, non sono rilevanti molte delle caratteristiche tipiche degli esseri umani, quali la personalità psichica e l'inclinazione alla malvagità<sup>307</sup>. Ciò che rileva è rappresen-

---

<sup>304</sup> F. Antolisei, *Manuale di Diritto Penale, Parte Generale*, cit. p. 240

<sup>305</sup> Si riporta per completezza l'intero testo dell'art. 41 c.p. "*Il concorso di cause preesistenti o simultanee o sopravvenute, anche se indipendenti dall'azione od omissione del colpevole, non esclude il rapporto di causalità fra l'azione od omissione e l'evento. Le cause sopravvenute escludono il rapporto di causalità quando sono state da sole sufficienti a determinare l'evento. In tal caso, se l'azione od omissione precedentemente commessa costituisce per sé un reato, si applica la pena per questo stabilita. Le disposizioni precedenti si applicano anche quando la causa preesistente o simultanea o sopravvenuta consiste nel fatto illecito altrui.*"

<sup>306</sup> G. Hallevy in, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit. 85 e ss.; lo stesso argomento è stato trattato sempre da G. Hallevy, *Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control*, cit. 171 e ss.

<sup>307</sup> Secondo la dottrina tedesca di Taterschuld sviluppatasi intorno agli anni 40, accanto alla colpa intesa in senso tradizionale si affiancherebbe la c.d. *colpa d'autore* avente ad oggetto le caratteristiche psichiche del reo e il suo stato soggettivo, determinando così un rimprovero all'individuo per la sua inclinazione al delitto

tato esclusivamente da due elementi: coscienza e volontà, quest'ultima declinabile secondo i livelli differenti di dolo, colpa o preterintenzione<sup>308</sup>. Si richiede, in altre parole, l'esistenza non solo di un nesso causale tra la condotta e l'evento ma altresì un nesso psichico tra l'agente ed il fatto, il quale si viene a creare tutte le volte che la condotta sia posta in essere coscientemente e volontariamente<sup>309</sup>.

## COSCIENZA

Per coscienza in particolare ci si riferisce alla rappresentazione da parte dell'agente di tutti gli elementi strutturali del fatto tipico, ovvero la condotta e la conseguente possibilità che l'evento si verifichi. L'agente dunque dovrà rappresentarsi l'eventualità che si verifichi un determinato avvenimento come risultato della propria condotta.

Più nello specifico il concetto di stato di coscienza, cd. Situation Awareness (SA), viene definito da Endsley, come *la percezione di elementi presenti nell'ambiente, in uno specifico volume di spazio e tempo, la comprensione del loro significato e la stima e la previsione dei loro stati futuri*<sup>310</sup>. Sulla base di tale nozione si possono distinguere tre livelli di coscienza:

- livello 1: percezione degli elementi che caratterizzano l'ambiente;
- livello 2: comprensione di tali elementi e della realtà in uno stato presente;
- livello 3: previsione degli stati futuri.

Nel primo capitolo della trattazione si sono analizzate le particolari modalità di funzionamento dei sistemi di A.I.<sup>311</sup> e, sulla base di queste, si può pacificamente constatare che i prodotti di machine learning appaiono in grado, attraverso i sensori di cui sono dotate<sup>312</sup>, di acquisire informazioni, spesso in modo anche più dettagliato rispetto alle capacità di un essere umano. Il primo dei tre livelli desumibili dalla definizione di Coscienza fornita da Endsley sembrerebbe pertanto soddisfatto.

---

e per la sua malvagità. Si tratta di una colpa che in definitiva giustificerebbe uno speciale aggravamento di pena. Tuttavia tale dottrina è stata aspramente criticata e non si ritiene condivisibile. Per un approfondimento si veda V. Bettiol, *Scritti giuridici*, Padova, Cedam, 1966, pp. 535-564.

<sup>308</sup> Così dispone l'art. 42. primo e secondo comma c.p. secondo il quale "*Nessuno può essere punito per un'azione od omissione preveduta dalla legge come reato, se non l'ha commessa con coscienza e volontà. Nessuno può essere punito per un fatto preveduto dalla legge come delitto, se non l'ha commesso con dolo, salvo i casi di delitto preterintenzionale o colposo espressamente preveduti dalla legge*"

<sup>309</sup> Vedi G. Fiandaca, *Diritto penale. Parte generale*, VII ed., Zannichelli, Roma, 2019, pp. 275 e ss.

<sup>310</sup> M. R. Endsley, D. J. Garald, *Theoretical underpinnings of situation awareness: A critical review*, in «Situation awareness analysis and measurement», 1, 2000, pp. 3-21.

<sup>311</sup> Cap. 1 par. relativo *modalità di funzionamento dei sistemi di intelligenza artificiale*.

<sup>312</sup> Si fa riferimento ai sensori di Radar, Lisar, Gliss, Gps.

Con riferimento al secondo passaggio che caratterizza il processo di coscienza, la quasi totalità delle tecnologie basate sull'intelligenza artificiale sono dotate di un *cervello artificiale* incorporato nell'hardware che opera tramite algoritmi di machine learning. Pertanto così come gli esseri umani creano, all'interno della propria mente, immagini della realtà esterna, i sistemi di intelligenza artificiale sono in grado di rappresentarsi un'immagine del mondo che li circonda attraverso l'assorbimento di dati acquisiti, l'integrazione di essi con le informazioni raccolte precedentemente, il loro successivo trasferimento alle unità di elaborazione ed infine la loro comparazione con i cd. *pattern*, schemi ricorrenti presenti in memoria<sup>313</sup>.

Sembra inoltre potersi configurare nei confronti dei sistemi di intelligenza artificiale anche il terzo livello del processo di Situation Awareness rappresentato dalla capacità di compiere previsioni sugli stati futuri. Un AI, sempre tramite i particolari algoritmi di machine learning tramite cui opera, appare in grado di calcolare le probabilità che si verifichino determinati esiti come conseguenza di determinate azioni fra loro alternative, di utilizzare queste informazioni come base di eventuali scelte e, di conseguenza, di assumere una scelta che appaia conforme e ragionevole rispetto agli obiettivi preposti.

In definitiva, secondo Hallevy, il processo di analisi dei sistemi d'intelligenza artificiale risulta equivalente ai corrispondenti processi umani: così come gli esseri umani attraverso vista, olfatto, tatto e udito ricevono informazioni e, attraverso la loro analisi, ne comprendono il significato, nei moderni sistemi d'intelligenza artificiale, lo stesso processo di analisi e comprensione della realtà, è svolto mediante le reti neurali artificiali e avanzati algoritmi di machine learning. Una volta raccolti e processati tutti gli elementi fattuali, il robot avrà quindi le informazioni sufficienti e necessarie per prendere una decisione cosciente e, di conseguenza, agire<sup>314</sup>.

---

<sup>313</sup> G.Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit., pp.53 e ss.

<sup>314</sup> G. Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, op.cit, pp. 53 e ss. In particolare Hallevy porta l'esempio di un Security Robot basato su un sistema d'intelligenza artificiale il cui compito è quello di identificare e segnalare alla polizia eventuali intrusi. L'identificazione è basata sugli elementi e i dati percepiti: l'aspetto fisico o il colore dei vestiti. Sulla base di essi il robot effettuerà un calcolo di tipo probabilistico e, nel caso in cui tale risultato non si dimostri sufficiente a produrre un'identificazione accurata, il robot inizierà allora un processo d'identificazione vocale, o ancora potrebbe richiedere al potenziale intruso un codice di identificazione o una password. Una volta acquisite le nuove informazioni, processerà tali dati e li integrerà con le informazioni già presenti in memoria, così da essere certo di non confondere gli intrusi con i membri delle forze di polizia o del personale di sicurezza.

## VOLONTÀ

Per quanto riguarda l'elemento della volontà, invece, non è necessariamente richiesto che l'azione o l'omissione che abbia causato l'evento sia stata volontaria, piuttosto è sufficiente che l'azione tragga origine o da un impulso volontario, ossia che il soggetto agente sia stato volontariamente attivo o inattivo, ovvero che l'azione derivi da un'inerzia del volere, la volontà ha infatti senza dubbio un campo che può dominare con i suoi poteri di impulso e inibizione e pertanto, assumendo una simile concezione della volontà, rientrerebbero in tale concetto anche tutti quegli atti che non sono preceduti da una chiara nozione di un fine, quali gli atti riflessi, gli atti istintivi e quelli abituali.<sup>315</sup>

I sistemi di machine learning, osserva Hallevy, hanno in effetti la capacità di soppesare tali fattori e, in alcuni casi, il processo di decision making può essere monitorato. Tuttavia, negli altri casi, la ricerca dell'elemento della volontarietà dell'azione appare più complesso. La volontarietà di un atto è in effetti uno stato interno all'agente e accertarlo con certezza non appare possibile, tuttavia, questo può essere dedotto esaminando il corso della condotta tenuta dal prodotto dell'intelligenza artificiale<sup>316</sup>.

Per la prova dell'intenzione, almeno per quanto riguarda i reati dolosi, Hallevy fa ricorso alla *Foreseeability rule presumption*. Si presume, nello specifico, che l'agente abbia voluto un determinato risultato se, mentre metteva coscientemente in atto la propria condotta, aveva previsto con una probabilità molto elevata il verificarsi di tale evento. Il riferimento è chiaramente rivolto alla figura del dolo indiretto. La ratio alla base di questa presunzione riposa sulla circostanza per cui l'agente, pur prevedendo con un elevato grado di probabilità il verificarsi dell'evento come risultato della propria condotta, decide di non modificarla e da ciò è possibile presumere che l'agente voglia effettivamente il verificarsi di un determinato accadimento<sup>317</sup>.

La prova della sussistenza della *mens rea* all'interno di un A.I. dovrà essere tuttavia valutata caso per caso, prendendo in considerazione le specifiche caratteristiche del singolo agente nella situazione concreta.

---

<sup>315</sup> F. Antolisei, *Manuale di Diritto Penale, Parte Generale*, cit., pp 341 e ss.

<sup>316</sup> G. Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit., pp 58 e ss.

<sup>317</sup> G. Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit., pp. 58 e ss.

Hallevy costruisce poi il medesimo schema, seppur adattandolo alle diverse forme di mens rea, per i reati colposi e per le ipotesi di *Strict liability*<sup>318</sup>, giungendo così alla conclusione che i moderni sistemi d'intelligenza artificiale sono in grado di soddisfare i requisiti sia dell'elemento oggettivo che dell'elemento soggettivo del reato.

L'agire dell'agente intelligente può qualificarsi infatti imprudente, *reckless*, nei casi in cui il sistema non abbia preso in considerazione una probabilità che avrebbe dovuto calcolare sulla base delle informazioni raccolte o già presenti nel sistema ovvero nell'eventualità in cui si verifichi un errore di calcolo nei processi di apprendimento, cd. *miscalculation*<sup>319</sup>.

Per l'autore dunque non vi sarebbe alcun motivo valido per non sottoporre i robot alla stessa responsabilità penale cui sono soggetti gli esseri umani nel caso in cui commettano un reato. In particolare, secondo il modello di responsabilità da lui costruito, l'incriminazione del prodotto dell'intelligenza artificiale non escluderebbe né sostituirebbe la responsabilità di utenti e programmatori nel caso in cui abbiano concorso con la propria condotta volontaria o negligente alla produzione dell'evento dannoso.

Hallevy propone in definitiva un modello di responsabilità per i prodotti dell'A.I. che si accosta, per alcuni versi, a quello dal Dlgs 231/2001, ma che se ne discosta per un aspetto decisamente non secondario: se nel quadro della responsabilità per gli enti il fatto di reato viene posto in essere da una persona fisica, apicale o subordinato, nel modello di responsabilità da lui proposto, lo evento lesivo viene posto in essere direttamente da un robot, un soggetto la cui coscienza e volontà si manifesta in maniera nettamente differente da quella umana.

## **IL PROBLEMA DELLE SANZIONI PER I PRODOTTI DELL'INTELLIGENZA ARTIFICIALE**

Come sopra accennato, Hallevy nel suo lavoro non si è limitato ad elaborare i tre diversi modelli di responsabilità applicabili ai sistemi di A.I., sostenendo ed appurando la possibilità che essi siano ritenuti penalmente responsabili per le azioni da loro poste in essere, ma è andato anche oltre: costruisce un articolato sistema di sanzioni applicabili ai robot,

---

<sup>318</sup> *Ivi*, pp. 106 e ss. Per la configurazione del delitto colposo in particolare, è richiesto dal nostro ordinamento che la condotta sia cosciente e volontaria, e che l'evento, salvo determinati casi, non sia voluto ma che sia imputabile all'agente per negligenza, imprudenza o imperizia.

<sup>319</sup> G.Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit., pp. 48 e ss.

sulla base dei medesimi fondamenti teorici che il diritto penale prevede per gli esseri umani.

Una futura ed eventuale teoria della responsabilità penale per i sistemi di intelligenza artificiale non dovrebbe, e non potrebbe, trascurare il profilo delle sanzioni comminabili a tali sistemi. Si tratta di una questione di rilevanza assolutamente centrale stante il fatto che il rilievo penale della responsabilità è sicuramente, in parte anche se non esclusivamente, connesso alla tipologia di sanzioni applicabili all'esito dell'accertamento della responsabilità<sup>320</sup>. Come ha stabilito la Corte EDU nella pronuncia *Engel e altri c. Paesi Bassi* infatti, al fine di stabilire la sussistenza di una *accusa in materia penale*, occorre tener presente tre criteri, noti come criteri di Engel, il terzo dei quali prende in considerazione la natura e il grado di severità della sanzione<sup>321</sup>.

Pertanto, al fine di elaborare un modello completo di responsabilità penale per i sistemi di intelligenza artificiale, alla questione *machina delinquere potest* occorrerà necessariamente affiancarvi il connesso quesito *machina puniri potest*. Ed è proprio questo l'intento perseguito da Hallevy nel suo lavoro *The Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control*<sup>322</sup>.

Secondo l'autore, le pene previste dagli ordinamenti penali ben si prestano ad essere applicate, seppur con i dovuti adattamenti ai prodotti dell'intelligenza artificiale, lasciando peraltro intatto, sia che si tratti di esseri umani, sia che si tratti di AI., il significato che la pena assume nel sistema penalistico.

Egli parte dal presupposto secondo cui, a differenza delle persone giuridiche, per le quali si è ora superato il paradigma *societas delinquere non potest*, i sistemi di intelligenza artificiale sono sempre dotate di un corpo fisico su cui la pena è in grado di incidere, in quanto, sebbene sia il software a elaborare informazioni necessarie per compiere una determinata azione, questa si estrinseca nel mondo esteriore tramite l'hardware, ossia la componente fisica della macchina.

---

<sup>320</sup> F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, cit., p. 31.

<sup>321</sup> Pronuncia Corte EDU, *Engel e altri c. Paesi Bassi*, 8 giugno 1976, §82, serie A n. 22. Per un commento della questione si veda Cass. pen, *Considerazioni sul principio del ne bis in idem nella recente giurisprudenza Europea: la sentenza del 4 Marzo 2014, Grande Stevens e altri contro Italia*, Rel. n. 35/2014, Roma 8 maggio 2014.

<sup>322</sup> La teoria prospettata dall'autore viene ribadita anche nei successivi lavori tra cui si cita G. Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit. pp. 156 e ss.

Sulla base di quanto esposto, Hallevy costruisce un sistema di equivalenze tra le pene attualmente previste per le persone fisiche e quelle che potrebbero o dovrebbero essere inflitte al robot.

Così, la pena della reclusione, che nei sistemi giuridici occidentali assume il significato di limitare o privare il soggetto della propria libertà personale, potrebbe essere tradotta nello spegnimento temporaneo, per un periodo di tempo limitato, dell'agente intelligente, privandolo così della libertà di agire nella propria area di pertinenza.<sup>323</sup>

Allo stesso modo, la condanna ai servizi sociali che nell'attuale ordinamento giuridico assume sostanzialmente una funzione risarcitoria rispetto al danno arrecato alla collettività risulterebbe in grado di svolgere il medesimo ruolo nei confronti di un A.I. qualora questo venga impiegato per svolgere lavori di pubblica utilità<sup>324</sup>.

Un ultimo riferimento infine viene fatto con riguardo alla pena capitale, che, se nel sistema di pene previsto per le persone fisiche è stata abrogata, nell'arsenale sanzionatorio previsto nei confronti degli enti torna in qualche modo in auge<sup>325</sup>. La cancellazione definitiva del programma d'intelligenza artificiale che controlla l'A.I., equivale alla pena di morte prevista per le persone fisiche o l'interdizione definitiva dall'esercizio dell'attività per le persone giuridiche: una volta terminato il programma, il robot sarà impossibilitato a commettere ulteriori reati. La cancellazione del programma, termina in definitiva, l'esistenza autonoma e indipendente dell'agente.<sup>326</sup>

La problematica maggiore, messa in evidenza già all'inizio del paragrafo, tuttavia riguarda la possibilità per la sanzione penale di svolgere, nei confronti dei sistemi di A.I., le stesse funzioni che essa è deputata a svolgere nei confronti delle persone fisiche.

Cercando di riassumere il pensiero di Hallevy, egli parte dalla concezione tradizionale secondo la quale la pena assolve generalmente tre funzioni: una funzione retributiva (in base alla quale la sanzione viene inflitta al fine di retribuire il male recato attraverso

---

<sup>323</sup> G.Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit. pp. 167 e ss.

<sup>324</sup> *Ivi*, pp. 171 e ss.

<sup>325</sup> Si fa riferimento al Dlgs 231/2001 art 9 c. 2 lett. a) "sanzione interdittiva dall'esercizio dell'attività" che, in qualche modo può essere equiparata alla pena capitale. Per una disamina sull'argomento si veda G. Lattanzi, *Reati e Responsabilità degli Enti, guida al dlgs. 8 giugno 2001, n.231*, Giuffrè Editore, Milano, 2010.

<sup>326</sup> G.Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit., p. 166.

l'azione criminosa), una funzione general-preventiva (in base alla quale la stessa minaccia della pena funge da deterrente nei confronti di possibili autori del reato) e una funzione special-preventiva (l'inflizione della pena gioca il ruolo di porre il condannato nella condizione di astenersi dal compiere, in futuro, ulteriori reati)<sup>327</sup>.

L'ultima delle tre funzioni, quella special-preventiva, ben potrebbe spiegare la propria efficacia anche nei confronti di un sistema di A.I., in quanto, è la stessa modalità di funzionamento del sistema di machine learning, analizzata nel primo capitolo, che permette alla macchina di imparare dall'esperienza e, in tal senso, una punizione nei confronti della stessa o un *training rieducativo*, potrebbe indurla ad astenersi dall'assumere in futuro una determinata condotta.

Più complicato risulta invece immaginare l'assolvimento di una funzione retributiva e general-preventiva della sanzione penale nei confronti di entità non umane.

Si potrebbero azzardare in tale sede ipotesi al momento poco più che fantascientifiche, quali la formulazione del precetto penale in termini digitali in modo tale da poter essere recepito ed elaborato dall'intero complesso dei sistemi di I.A., ovvero la trasmissione dell'esperienza della pena subita dal singolo computer anche ai computer consimili, futuri potenziali autori di reati<sup>328</sup>.

## CONCLUSIONI

In definitiva si può constatare che la disciplina della responsabilità penale nei confronti delle persone giuridiche introdotta nel 2001 ha certamente costituito una svolta innovativa nell'area del diritto penale e, i modelli da essa prospettati possono costituire preziosi indizi al fine di definire un eventuale e futuro quadro normativo volto a disciplinare la responsabilità penale degli agenti artificiali.

Il diritto penale infatti, contrariamente a quanto possa immaginarsi a un primo sguardo, essendo circondato da numerose garanzie costituzionali e sovranazionali, dimostra un certo grado di flessibilità qualora ciò risulti necessario per ragioni di politica criminale, alla obbligata condizione che certe premesse dogmatiche siano in ogni caso soddisfatte.

---

<sup>327</sup> Per un approfondimento sulle "teorie della pena", D. Pulitanò, *Diritto penale*, Giappichelli, Torino, 2017, pp. 47 ss.; e F. Palazzo, *Corso di diritto penale*, cit., pp. 16 ss.

<sup>328</sup> F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, cit., pp. 32.

## OPINIONI CONTRARIE ALL'ATTRIBUZIONE DELLA RESPONSABILITÀ DIRETTAMENTE AL ROBOT

La tesi sostenuta da Hallevy è senza dubbio accompagnata da una forte carica innovativa, ma allo stesso tempo, presta il fianco a numerose obiezioni poggiate in parte su pregiudizi di natura antropocentrica, ma prima ancora su considerazioni relative ai principi fondamentali del diritto penale. Inoltre, come prima evidenziato, accennare a un parallelismo tra la disciplina relativa alla responsabilità penale degli enti e quella, ipotetica, dei robot non appare del tutto scontato stante le sostanziali differenze esistenti tra tali entità.

Va precisato fin da subito che la proposta del Parlamento Europeo del febbraio 2017 relativa alla creazione di uno status giuridico ad hoc per i robot, o almeno per quelli più evoluti, è stata fortemente e formalmente criticata. Sebbene, come rilevato più volte nel corso della trattazione, la proposta faccia esclusivo riferimento all'ambito civilistico, il riconoscimento di una *soggettività elettronica* rappresenta il presupposto della teoria sostenuta da Hallevy. Non si può dunque prescindere dal prendere in considerazione e analizzare brevemente le motivazioni poste alla base di tali critiche.

In particolare, l'eurodeputata e vicepresidente della commissione Affari legali del Parlamento europeo, Mady Delvaux, e altri 156 esperti di intelligenza artificiale, tra cui informatici, professori delle facoltà di giurisprudenza e amministratori delegati di varie società, provenienti da 14 paesi europei, avvertono l'attribuzione di una personalità legale ai prodotti dell'intelligenza artificiale come *una scelta inappropriata sia da una prospettiva legale che etica*<sup>329</sup>.

La lettera parte dal presupposto meramente tecnico secondo cui una scelta in tale direzione sarebbe fondata su una fuorviante sopravvalutazione delle attuali capacità dei robot, compresi quelli maggiormente evoluti, essendo ciò dovuto alla periodica pubblicazione di articoli e comunicati stampa volti a esasperare il reale stato della tecnica.

Si sostiene poi che, anche da un punto di vista legale ed etico, l'attribuzione al robot di una personalità giuridica sarebbe inappropriata in quanto:

- questa non potrebbe essere derivata dal *modello di persona fisica*<sup>330</sup> in quanto così facendo si attribuirebbe al prodotto dell'intelligenza artificiale tutti i più fondamentali diritti umani, quale il diritto all'integrità, alla dignità, alla cittadinanza e alla retribuzione. Questo sarebbe contrario a quanto previsto dalla Carta Europea dei Diritti

---

<sup>329</sup> *Open Letter to the European Commission Artificial Intelligence and Robotics*, 05/04/2018 <http://g8fip1kplyr33r3krz5b97d1-wpengine.netdna-ssl.com/wp-content/uploads/2018/04/RoboticsOpen-Letter.pdf>

<sup>330</sup> *Ivi*, in particolare si fa riferimento al "Natural Person Model".

dell'uomo nonché alle diverse Convenzioni per la protezione dei diritti umani e delle libertà fondamentali.

- non potrebbe altresì essere desunta dal *modello legale di identità*<sup>331</sup>, in quanto ciò implicherebbe l'esistenza di una persona umana dietro la persona legale che la rappresenti e la diriga. E questo non è il caso dei robot, ma si tratterebbe piuttosto di una mera *fictio iuris*<sup>332</sup>.

Tornando alla tesi sostenuta da Hallevy, al riconoscimento dei sistemi di intelligenza artificiale come soggetti attivi del reato viene obiettata principalmente la mancanza di imputabilità nei confronti di tali agenti, anch'essa data quale presupposto della teoria sulla responsabilità da lui esposta.

L'imputabilità, così come disciplinata dall'articolo 85 c.p. richiede la sussistenza, in capo al soggetto attivo del reato, della *capacità d'intendere e di volere*<sup>333</sup>.

Per capacità di volere si intende, più nello specifico, la capacità di determinarsi in modo autonomo resistendo agli impulsi mentre per capacità di intendere si fa riferimento alla capacità di rendersi conto del valore sociale dell'azione che si compie, ossia che l'azione posta in essere contrasta con le esigenze di vita in comune<sup>334</sup>.

Alla base di tale obiezione riposa la circostanza per cui, anche ammettendo l'insegnamento di Hallevy per cui l'elemento della colpevolezza risulta provato in capo al robot, lo stesso non può dirsi per quello relativo all'imputabilità: la coscienza e la volontà richieste per la prova della colpevolezza vanno tenute distinte dalla capacità di intendere e di volere intese quali capacità o presupposto dell'imputabilità<sup>335</sup>. In tal senso è significativa la circostanza per cui l'articolo 85 viene collocato dal Codice Rocco non nella parte relativa alla colpevolezza e all'elemento soggettivo del reato, ma piuttosto, nella parte relativa al reo. È necessario pertanto separare la colpevolezza dall'imputabilità: la prova

---

<sup>331</sup> *Ivi*. Si fa in tal caso riferimento al "Legal Entity model"

<sup>332</sup> *Ivi*.

<sup>333</sup> Il testo completo dell'articolo recita "Nessuno può essere punito per un fatto preveduto dalla legge come reato, se, al momento in cui lo ha commesso, non era imputabile. È imputabile chi ha la capacità d'intendere e di volere".

<sup>334</sup> F. Antolisei, *Manuale di Diritto Penale, Parte Generale*, cit., pp. 609 e ss.

<sup>335</sup> Sull'argomento si rinvia a M. Romano, G. Grasso, *Commentario Sistematico del Codice Penale*, IV ed., vol. II, Giuffrè, Milano, 2012, pp. 31 e ss., che rimanda a M. Spasari, *Diritto criminale e scienza del diritto criminale, Oggetto metodo della dogmatica criminale e della criminologia*, in «Giurisprudenza italiana», 4, 1990, pp. 506 e ss.

della sussistenza in capo all'agente intelligente del dolo o della colpa non implica necessariamente l'imputabilità dello stesso<sup>336</sup>.

L'articolo 42 detta pertanto le condizioni necessarie affinché un determinato atto possa essere attribuito alla volontà dell'agente, diversamente l'articolo 85 disciplina le qualità richieste affinché un soggetto possa essere assoggettato a pena e, nel contesto in esame, non sembra che queste possano dirsi sussistenti. Come prima accennato, la capacità di intendere viene definita come la capacità di rendersi conto del valore sociale dell'azione che si compie. Si tratta tuttavia di un intendere *mediato*: è il soggetto umano che opera dietro al software che immette tale valore sociale nella memoria dell'agente artificiale.

Il fatto che un robot accordi un valore maggiore alla vita di un essere vivente piuttosto che all'integrità di un oggetto inanimato non è dovuto all'autonoma capacità dello stesso di intenderne l'intrinseco valore. Per il robot, sia esso un'automobile driverless o un drone militare, la persona o l'oggetto appariranno come semplici ostacoli all'obbiettivo da raggiungere; ad esempio i sistemi di radar e lidar, paragonabili agli occhi di un automobilista umano, percepiranno esclusivamente una massa e la propria distanza da essa e, di conseguenza, valuteranno la possibilità di aggirarla o ignorarla, ma ciò in base a fattori puramente tecnici e oggettivi.

Sarà il programmatore a introdurre i dati e le informazioni necessarie, accompagnate da un intenso training, affinché la macchina possa imparare a comprendere il valore sociale delle azioni compiute.

Riprendendo, ancora una volta, l'insegnamento del Professor Floridi non appare dunque appropriato accostare la parola intelligenza con artificiale: la prima è e rimane una qualità che appartiene esclusivamente e gelosamente all'essere vivente<sup>337</sup>. Bisogna tener sempre distinti questi concetti anche nel momento in cui si ragiona in termini di responsabilità penale degli stessi; il fatto che la tecnologia abbia permesso alla macchina di raggiungere lo stesso risultato di un essere umano non significa che il processo adottato e la fonte, ossia l'intelligenza, siano le stesse<sup>338</sup>. Non sono i prodotti dell'intelligenza artificiale che

---

<sup>336</sup> Vedi per tutti F. Antolisei, *Manuale di Diritto Penale, Parte Generale*, cit., pp. 326 e ss.

<sup>337</sup> L. Floridi, *La vita al tempo del digitale*, cit.

<sup>338</sup> Sempre L. Floridi, *La quarta rivoluzione*, cit. pp. 165 e ss.

stanno diventando sempre più intelligenti, ma è il mondo, al contrario, che si sta trasformando in un *'infosfera*<sup>339</sup> sempre più adatta e accogliente alle limitate capacità di tali macchine. *“Se immergiamo gli agenti artificiali nella loro zuppa digitale li ritroveremo felicemente ronzanti”*<sup>340</sup>.

Sulle stesse motivazioni si fonda la critica per cui, in realtà, nemmeno l'elemento della colpevolezza si potrebbe ritenere sussistente in capo al robot, come invece sostenuto e analizzato da Hallevy. La colpevolezza, intesa quale coscienza e volontà di un'azione<sup>341</sup> postula invero una libertà del volere; libertà di cui, per le stesse ragioni esposte con riferimento alla capacità di intendere richiesta ai fini dell'imputabilità, il prodotto dell'intelligenza artificiale risulta privo. Fino a che punto è possibile affermare che la discrezionalità dell'uomo sia paragonabile a quella di una macchina?<sup>342</sup>

Sembrerebbe prospettarsi così il dibattito, ancora aperto, sul *libero arbitrio*, inteso come concetto filosofico e teologico secondo il quale ogni persona avrebbe la facoltà di scegliere il fine del proprio agire e pensare, tipicamente perseguiti tramite volontà. La possibilità di scelta avrebbe così origine direttamente dalla e nella persona stessa e non sarebbe determinata da forze esterne. Sulla base di ciò un soggetto sarebbe meritevole di sanzione penale in quanto causa cosciente e libera del fatto commesso<sup>343</sup>. Ammettendo quindi l'esistenza del libero arbitrio non si può non constatare che l'agente artificiale ne sarebbe privo in quanto, anche i robot più avanzati e autonomi sono stati programmati per agire in un determinato modo, l'azione degli stessi sarebbe sempre determinata da forze esterne.

L'obiezione più critica riguarda infine la possibilità di sottoporre a sanzione penale il robot. La pena, si ripete, nell'ordinamento penale svolge essenzialmente tre funzioni: retributiva, general-preventiva e special-preventiva. Solo l'ultima di esse tuttavia sembrerebbe essere compatibile con un'eventuale punizione del robot.

---

<sup>339</sup> Il termine viene utilizzato da Luciano Floridi per indicare la graduale informatizzazione dell'intero sistema sociale, una società sempre più basata e condizionata dalle informazioni a disposizione. Per un approfondimento sul tema si veda L. Floridi, *La quarta rivoluzione*, cit., pp. 27 e ss.

<sup>340</sup> L. Floridi, *La quarta rivoluzione*, cit. p. 155.

<sup>341</sup> L'elemento si ricava, come esposto precedentemente, dall'articolo 27 Cost. e 42 c.p.

<sup>342</sup> R. Borsari, *Intelligenza artificiale e responsabilità penale: prime considerazioni*, in «Rivista di diritto dei media», 3, 2019, p. 262.

<sup>343</sup> Tra i sostenitori di tale teoria si menzionano L. Scarano, *Libera volontà e libero Arbitrio nel Diritto Penale*, Giuffrè, Milano, 1937 e G. Bettiol e L. Pettoello Mantovani, *Diritto Penale*, cit., pp. 459 e ss.

Si potrebbe certo ipotizzare, come sopra accennato, la formulazione del precetto penale in termini digitali in modo tale da poter essere recepito ed elaborato anche dai sistemi di I.A. nel loro complesso, ovvero la trasmissione dell'esperienza della pena subita dal singolo computer anche ai computer consimili, futuri potenziali autori di reati<sup>344</sup>.

Si tratta tuttavia di ipotesi, allo stato, poco concrete e difficilmente attuabili ma che, se ci fossero le possibilità e le competenze tecniche per svilupparle, sarebbero in grado permettere alla sanzione penale di assolvere nei confronti di tali soggetti anche la sua funzione general-preventiva.

L'assolvimento della funzione retributiva della pena, in definitiva, appare pertanto l'ostacolo maggiore che si oppone al riconoscimento di una responsabilità penale in capo agli agenti artificiali. Secondo le teorie della retribuzione la pena altro non è che una ricompensa: *malum passionis quod infligitur ob malum actionis*, secondo la celebre definizione di Ugo Grozio<sup>345</sup>. Il reo merita un castigo, deve essere punito. Il criterio generale citato ha poi assunto diversi atteggiamenti, cui i principali si identificano come: *retribuzione giuridica e retribuzione morale*.

Dal punto di vista giuridico, la funzione retributiva assolve lo scopo di ripristinare la legalità violata; la pena viene quindi giustificata per una ragione esclusivamente interna all'ordinamento giuridico, il quale dovrebbe garantire e tutelare se stesso punendo coloro che non ne seguono i precetti. Hegel ha fornito a tale dottrina una forma dialettica: se il delitto costituisce la negazione del diritto e la pena rappresenta, a sua volta, la negazione del delitto si sarebbe allora di fronte a una negazione di una negazione; la pena riaffermerebbe così il diritto<sup>346</sup>. Ci si chiede tuttavia se il diritto possa effettivamente dirsi riaffermato infliggendo un castigo alla macchina il cui meccanismo di funzionamento si è evoluto, per ragioni non sempre conosciute e conoscibili, in una direzione diversa da quella prevista.

La teoria della responsabilità sostenuta da Hallevy, tuttavia, incontra il suo principale limite se posta dinnanzi alla *teoria della retribuzione morale*. Essa si fonda sull'esigenza

---

<sup>344</sup> Così F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, cit., p. 32.

<sup>345</sup> Trad. "Il male della sofferenza è inflitto per il male dell'azione"

<sup>346</sup> Hegel, *Lineamenti di filosofia del diritto*, trad. it. a cura di F. Messineo, Laterza, Bari, 1913, pp. 92 e ss.

profonda e incoercibile della natura umana che il male venga retribuito col male. La retribuzione del colpevole rappresenterebbe così un imperativo di giustizia, un'esigenza che trova in sé la sua giustificazione, senza la necessità di essere ricercata in utilità esterne.

La dottrina in esame ha assunto la sua più alta espressione con Kant, il quale qualifica la legge penale come un imperativo categorico. Scrive il filosofo a tal proposito che *Anche qualora la società civile si sciogliesse con il consenso di tutti i suoi membri, l'ultimo assassino che si trovasse in prigione dovrebbe essere giustiziato, affinché ciascuno porti la pena della sua condotta e il sangue non ricada sul popolo che non ha reclamato la punizione*<sup>347</sup>.

Appare tuttavia difficile giustificare l'assunto per cui il male debba essere retribuito con il male nel caso in cui esso venga inflitto a un soggetto che sia privo di quell'emotività e quella sensibilità che caratterizza gli esseri umani. Non solo, seguendo l'insegnamento di tale teoria, poiché il delitto costituisce una violazione dell'ordine etico, la coscienza morale ne esigerebbe una punizione. Ma è veramente possibile affermare che la coscienza morale sia ripagata da una punizione inflitta al robot?

Tali ostacoli sono i medesimi che si sono presentati, e sono stati successivamente risolti, con riguardo all'introduzione della disciplina della responsabilità penale per gli enti. Si potrebbe allora pensare che alla base di tali critiche altro non risieda che un semplice pregiudizio di natura antropocentrica ma, osservando più nel dettaglio la disciplina emerge che in realtà la questione si presenta più complessa. Nel sistema normativo previsto dal dlgs 231/2001 l'illecito dell'ente è connesso al reato posto in essere dal subordinato o dall'apicale, quindi al reato della persona fisica: vengono puniti, anche se assoggettati a sanzioni differenti, sia la persona fisica che quella giuridica<sup>348</sup>. Al contrario nella disciplina proposta da Hallevy il reato del robot può anche prescindere dal reato della persona fisica che opera dietro di esso. Il prodotto dell'intelligenza artificiale sarebbe dunque l'unico soggetto a subire gli effetti della sanzione penale, dando così luogo ai dubbi e alle problematiche sopra esposte.

---

<sup>347</sup> Così Kant, *La dottrina del diritto*, trad. it. con prefazione e note di Giovanni Vidari, Studio editoriale lombardo, Milano, 1916, pp. 144 e ss.

<sup>348</sup> Per un approfondimento Cfr. G. Lattanzi, *Reati e Responsabilità degli Enti, guida al dlgs. 8 giugno 2001, n.231*, cit. pp. 56 e ss.

## 2.4. Una seconda proposta: la responsabilità del produttore e del programmatore

Si analizza ora la diversa ipotesi sostenuta in ambito penalistico dalla dottrina maggioritaria: quella di estendere la categoria penalistica della responsabilità da prodotto difettoso anche ai danni provocati da un'evoluzione anomala del software degli agenti artificiali.

Nella stessa direzione si muove anche la proposta del Parlamento Europeo contenuta nella Risoluzione del febbraio 2017 relativa, come accennato più volte, alla sola responsabilità civile relativa ai danni causati dai prodotti dell'A.I.

Il Parlamento Europeo ha sostenuto infatti di estendere, interpretando in maniera evolutiva, le disposizioni già esistenti nell'ordinamento comunitario con particolare riferimento alla direttiva 85/374/CEE relativa al *ravvicinamento delle disposizioni legislative, regolamentari ed amministrative degli Stati Membri in materia di responsabilità per danno da prodotti difettosi*.

Tali considerazioni poggiano sull'idea secondo cui l'autonomia del robot non parrebbe una caratteristica di per sé idonea a giustificare l'elaborazione di nuove categorie giuridiche o di nuove norme, ma sarebbe piuttosto sufficiente un'interpretazione evolutiva delle disposizioni attualmente esistenti in materia di responsabilità da prodotto difettoso<sup>349</sup>. Più nello specifico, il Parlamento Europeo prospetta due possibili approcci al fine di inquadrare la responsabilità dell'*uomo dietro la macchina* per i danni da essa provocati:

- la prima è rappresentata da un'ipotesi di *Strict liability*, ossia di responsabilità oggettiva per la quale è richiesta esclusivamente la prova del danno avvenuto e l'individuazione del nesso di causalità che collega l'azione lesiva posta in essere dal robot al danno subito dalla parte lesa<sup>350</sup>.

---

<sup>349</sup> U. Ruffolo, *Intelligenza artificiale e responsabilità*, cit., p. 84.

<sup>350</sup> Parlamento europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, cit. pt. 54.

- la seconda consiste piuttosto in un approccio di gestione dei rischi. L'attenzione ricade in questo caso sul soggetto in grado di minimizzare i rischi e affrontarne l'impatto negativo<sup>351</sup>.

Sebbene tali approcci siano presentati nella risoluzione come alternativi in realtà, va da subito precisato che l'uno non esclude l'altro in quanto l'approccio di gestione dei rischi non implica la necessaria sussistenza di alcun elemento di colpevolezza in capo ai soggetti individuati come responsabili<sup>352</sup>.

Come si è visto le problematiche presentate in ambito civilistico sono destinate ad ampliarsi qualora proiettate sul piano penale, soprattutto nel momento in cui vengono in rilievo ipotesi di responsabilità oggettiva. Se infatti la disciplina relativa alla responsabilità civile per danno da prodotto difettoso addebita l'evento lesivo sulla scorta della sola individuazione del nesso di causalità che collega il difetto del prodotto al danno subito dalla parte lesa, e se, in ambito penale la giurisprudenza, per muoversi nella stessa direzione, ha assunto un atteggiamento teso a forzare la tipicità sia oggettiva che soprattutto soggettiva del fatto illecito<sup>353</sup>, nel momento in cui vengano in rilievo i prodotti dell'A.I., il cui meccanismo di funzionamento appare per sua natura non del tutto prevedibile ex ante, la pedissequa osservanza di quei criteri pervicacemente applicati dalla giurisprudenza, sembrerebbe risolversi in un'inaccettabile forzatura dei principi fondamentali operanti in materia penale.

Si tenterà pertanto di prendere in considerazione tutte le diverse problematiche sollevate da un simile quesito, introducendo brevemente le questioni sollevate dalla letteratura con riferimento all'ipotesi di responsabilità dolosa e soffermandoci, più nello specifico, sulla più complessa ipotesi relativa alla responsabilità colposa dell'*uomo dietro la macchina*.

Di più agevole soluzione è sicuramente l'ipotesi dolosa, in quanto, in tali casi l'uomo resterebbe responsabile per le azioni da lui volute; il robot altro non rappresenterebbe che una *longa manus* di colui che ha programmato l'algoritmo. Certo, date le peculiarità di

---

<sup>351</sup> *Ivi*, pt. 55.

<sup>352</sup> Ugo Ruffolo, *Intelligenza artificiale e responsabilità*, cit., p. 87.

<sup>353</sup> C. Piergallini, *Danno da prodotto e responsabilità penale*, cit., pp. 440 ss., nonché più in breve C. Piergallini, *La responsabilità del produttore: una nuova frontiera del diritto penale?*, in «Diritto Penale e Processo», 9, 2007, pp. 1128 e ss.

tali sistemi sarebbe opportuno, in omaggio al principio di legalità, introdurre nuove categorie di delitti o quantomeno adattare e modificare la normativa vigente<sup>354</sup>. Si fa riferimento nello specifico all'articolo 111 c.p. che disciplina l'ipotesi in cui una persona non imputabile o non punibile venga determinata a commettere un reato, prevedendo in tali situazioni, la punibilità di colui che si serva di tale persona come mero strumento per commettere un'azione criminosa.

Tuttavia, dato l'elevato grado di autonomia e le capacità di *decision making* dell'agente artificiale, questo, come sottolineato anche in sede Europea, non può essere considerato alla stregua di un mero strumento ma, dall'altro lato, non potrebbe nemmeno essere ricompreso nel concetto di persona: sarebbe a tal proposito auspicabile una modifica dell'articolo 111 nella parte in cui fa esclusivo riferimento alla *persona non imputabile o non punibile*, in modo tale da estenderne l'ambito di applicazione anche nel caso in cui il prodotto dell'intelligenza artificiale venga intenzionalmente programmato in modo tale da commettere azioni criminose<sup>355</sup>.

L'ipotesi colposa solleva al contrario numerose problematiche di carattere non solo giuridico ma prima ancora etico: Quali sono le condizioni affinché l'uomo dietro la macchina possa essere ritenuto responsabile per le azioni poste in essere dal robot? A che titolo può essere imputata la responsabilità dell'uomo se si assume che l'agente artificiale sia in grado di assumere decisioni in modo autonomo? E infine, a chi tra il programmatore, il produttore o l'utilizzatore dovrà essere imputata la responsabilità?

#### **L'ESTENSIBILITÀ DELLA CATEGORIA PENALISTICA PER DANNO DA PRODOTTO.**

La tesi più ardentemente sostenuta in dottrina fa leva sulla possibilità di inquadrare la responsabilità dei diversi soggetti coinvolti nella catena produttiva nella categoria penalistica della responsabilità per danno da prodotto. Il riferimento va in particolare alla responsabilità dei programmatori, progettisti e costruttori del mezzo.

Si precisa nuovamente che l'attuale normativa relativa al danno da prodotto difettoso opera con esclusivo riferimento alle ipotesi di responsabilità civile ma, intorno ad essa, si

---

<sup>354</sup> M.B.Magro, *A.I.: la responsabilità penale per la progettazione, la costruzione e l'uso dei robot*, in «Pluris, quotidiano giuridico», 12 giugno 2018, p. 1.

<sup>355</sup> M. Iaselli, *Risoluzione UE sulla responsabilità dei robot: sarà mai possibile equiparare il robot ad un essere umano?*, in «ANDIP, associazione nazionale per la difesa della Privacy», 14 febbraio 2018.

è sviluppata una numerosa giurisprudenza e una cospicua letteratura volte a indagarne i risvolti penalistici e le relative elevatissime problematicità<sup>356</sup>.

Si noti tuttavia come, a differenza delle tipologie di prodotto solitamente prese in considerazione nell'analisi della responsabilità penale per danno da prodotto, il *prodotto intelligente* oggetto della suddetta analisi, appaia indiscutibilmente più complesso e contornato da ancora più numerosi profili di problematicità<sup>357</sup>.

Si è già discusso nel secondo capitolo della trattazione, circa la difficoltà di ricondurre il danno da prodotto di machine learning nelle ordinarie categorie di danno prodotto e dell'opportunità, se non altro nel lungo periodo, di affiancare a quelle preesistenti una nuova categoria che prenda specificatamente in considerazione i danni provocati da quei prodotti in grado di auto apprendere tramite l'esperienza. Le *self-driving cars* infatti, al pari di tutti i prodotti dell'intelligenza artificiale, non si identificano alla stregua dei tradizionali prodotti-oggetto passivi ma, almeno parzialmente, sono riconducibili a soggetti, proattivi, dotati di autonomia d'azione.

Dunque, la difficoltà di inquadrare i danni provocati dagli A.I nelle tradizionali categorie di danno da prodotto viene generalmente ricondotta non tanto a eventuali errori generatisi nella fase di progettazione e assemblaggio del mezzo, fasi che tutto sommato non si distaccano nettamente rispetto a quanto già oggi avviene con riferimento ai tradizionali prodotti, quanto piuttosto a errori che traggono origine nella fase della programmazione<sup>358</sup>.

La quantità di scenari che i programmatori sono chiamati a valutare è infatti vastissima e potenzialmente illimitata comportando di conseguenza l'intrecciarsi del lavoro di centinaia di informatici nella creazione di un prodotto finale rispetto al quale risulta molto

---

<sup>356</sup>Il riferimento principale va all'ampia monografia di C.Piergallini, *Danno da prodotto e responsabilità penale*, cit. ma si vedano già i suoi precedenti scritti tra cui si cita: C.Piergallini, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, in «Rivista Italiana di Diritto e Procedura Penale», 1996, pp. 354-370. e C. Piergallini, *Attività produttive e imputazione per colpa: prove tecniche di «diritto penale del rischio»*, in «Rivista Italiana di Diritto e Procedura Penale», 1997, pp. 1473-1495, oltre al più recente, efficacissimo sunto della questione in C. Piergallini, *La responsabilità del produttore: una nuova frontiera del diritto penale?*, cit. Sul tema cfr. altresì D. Castronuovo, *Responsabilità da prodotto e struttura del fatto colposo*, in «Rivista Italiana di Diritto e Procedura Penale», 1, 2005, 301-340.

<sup>357</sup> Cfr., S. Gless, E.Silverman, T. Weigend, *If robots cause harm, who is to blame? Self-driving cars and criminal liability*, in «New Criminal Law Review», 19, 2016, pp. 426 ss.

<sup>358</sup> A.Cappellini, *Profili penalistici delle self-driving cars*, cit., p. 338.

spesso arduo individuare lo specifico momento in cui l'errore ha avuto origine e il soggetto che ha contribuito alla generazione dello stesso.

Come ripetuto più volte nel corso della trattazione, le particolari modalità di machine learning tramite cui il prodotto opera e apprende nel tempo, non si fondano su istruzioni precise e specifiche per ogni tipo di problema al presentarsi del quale il robot non deve far altro che eseguire, ma al contrario prevedono una costante evoluzione delle capacità di azione del robot implementate sulla base delle informazioni originariamente inserite nel sistema stesso. A conferma di ciò basti pensare al fatto che gli algoritmi complessi attraverso cui il prodotto intelligente opera vengono spesso definiti *black box* proprio per indicare quell'opacità dei processi decisionali sviluppati internamente alla macchina<sup>359</sup>. Caratteristica questa che indubbiamente rende infinitamente più complessa la ricostruzione della trama causale del comportamento della macchina, nonché la misura della diligenza richiesta al produttore, elementi su cui si fonda l'addebito della responsabilità penale per danno da prodotto.

Si consideri infine che, poiché non tutti gli scenari possono essere anticipati e tradotti in istruzioni tramite algoritmi, residua un'ineliminabile componente di imprevedibilità nel comportamento degli A.I. Anzi, si potrebbe dire che, «*in una certa misura, l'imprevedibilità dell'agente intelligente è "pre-programmata", e con essa i rischi associati a terzi*»<sup>360</sup>.

Sulla base delle considerazioni ora svolte, non dovrà pertanto stupire l'estrema complessità, se non addirittura l'impossibilità, nell'individuare se vi sia effettivamente un responsabile: un soggetto che non abbia previsto un evento che avrebbe ragionevolmente dovuto prevedere, ovvero che abbia mal implementato nel sistema le modalità con cui il mezzo stesso dovrebbe reagire al presentarsi di specifici dati concreti.

Nei casi in cui l'errore tecnico di programmazione possa essere rilevato ed individuato con certezza, si ritiene che possano operare i tradizionali meccanismi individuati dalla giurisprudenza per imputare il danno da prodotto al soggetto responsabile dell'area di rischio in cui il difetto ha avuto origine<sup>361</sup>.

---

<sup>359</sup> *Ivi.*, pp. 339 e ss.

<sup>360</sup> M.B. Magro, A.I.: *la responsabilità penale per la progettazione, la costruzione e l'uso dei robot*, cit., p. 2.

<sup>361</sup> A.Cappellini, *Profili penalistici delle self-driving cars*, cit., p. 338.

Si è già analizzato nel secondo capitolo della trattazione come l'introduzione, in fase di addestramento del prodotto di *machine learning*, di un set di dati incompleto o pregiudizievole possa incidere sul comportamento futuro tenuto dalla macchina. Si tratta questa di un'ipotesi in cui il difetto che ha dato origine al comportamento lesivo dell'A.I. potrebbe essere individuato con un margine, più o meno ampio, di certezza. La responsabilità ricadrà in tali casi sul soggetto che ha mal governato il rischio che il processo decisionale della macchina potesse essere condizionato da distorsioni o pregiudizi.

Un'altra situazione in cui si rileva l'operatività dei detti criteri giurisprudenziali analizzati dalla letteratura di Piergallini riguarda il caso in cui i vertici aziendali, dopo essere venuti a conoscenza di un problema di malfunzionamento del software, non abbiano immediatamente provveduto a porvi rimedio: o imponendo un aggiornamento obbligatorio del sistema di guida autonoma che sani le criticità rilevate, oppure, come *extrema ratio*, ritirando il prodotto dal mercato. In tali ipotesi si passerebbe oltre il problema dell'individuazione del singolo programmatore responsabile ma si addebiterebbe il fatto direttamente al soggetto che riveste la posizione di vertice all'interno dell'unità organizzativa. Si tratta in ogni caso di meccanismi imputativi delineati e seguiti da una costante giurisprudenza ma che, come rilevato più volte a livello dottrinale, tendono indebitamente a forzare la tipicità sia oggettiva che soggettiva dell'illecito colposo d'evento<sup>362</sup>.

Tra i diversi problemi che insorgono sul terreno della responsabilità penale per danno da prodotto infatti, la letteratura di Piergallini individua, oltre alla difficoltà di un effettivo accertamento causale del rischio, problema che nella specifica ipotesi presa in considerazione si considera risolto, anche le difficoltà relative alla *ricostruzione della struttura del tipo* e difficoltà sul *piano della colpevolezza*<sup>363</sup>.

Per quanto riguarda il primo dei due aspetti problematici la giurisprudenza solitamente struttura l'illecito in questione sul modello del reato omissivo improprio<sup>364</sup>, focalizzando così l'attenzione sulla posizione di garanzia rivestita dai diversi soggetti coinvolti nell'apparato produttivo industriale. Si tratta tuttavia di una *posizione di garanzia da ingerenza*

---

<sup>362</sup> Sempre C.Piergallini, *Danno da prodotto e responsabilità penale*, cit., pp. 440 ss., nonché più in breve Piergallini, *La responsabilità del produttore: una nuova frontiera del diritto penale?*, cit. p. 1128.

<sup>363</sup> C.Piergallini, *Danno da prodotto e responsabilità penale*, cit., pp. 95-103.

<sup>364</sup> L'articolo 40 del codice penale al 2 comma detta la particolare disciplina dei reati omissivi c.d. impropri, per cui "non impedire un evento che si ha l'obbligo giuridico di impedire, equivale a cagionarlo".

per cui la responsabilità omissiva discende dalla semplice produzione e messa in circolazione di prodotti-fonte di pericolo da cui scaturisce il conseguente obbligo di intervenire sugli stessi tramite il loro ritiro dal mercato qualora si percepisca, anche lontanamente, la loro non completa sicurezza<sup>365</sup>.

Ma le problematiche si riflettono anche e soprattutto sul piano della colpevolezza. Si assiste in particolare a un forte indebolimento, se non a un vero e proprio processo di obiettivizzazione dell'elemento del dolo, che si riduce a una mera intuizione del rischio del verificarsi dell'evento lesivo<sup>366</sup>.

Per quanto riguarda infine l'elemento della colpa, si consideri che la difficoltà di formulare *ex ante* giudizi di prevedibilità ed evitabilità rende particolarmente ardua l'individuazione del grado della diligenza dovuta, tanto da far sorgere l'interrogativo se tali criteri non rappresentino effettivamente una *maschera a una forma non codificata di responsabilità oggettiva*<sup>367</sup>. Si consideri infine che, se nel normotipo classico la distinzione tra dolo e colpa risulta essere netta o quantomeno identificabile, nel paradigma della responsabilità penale per danno da prodotto la distinzione tra i due elementi risulta particolarmente difficoltosa in quanto spesso ci si imbatte in *decisioni in stato di incertezza*, che muovono da una piattaforma rappresentativa molto indebolita, capace di tramandare la fisionomia di un *rischio meramente intuibile* più che un rischio *scientificamente apprezzabile*.

Il nodo realmente problematico attinente alla materia in esame si presenta tuttavia nei casi in cui l'evento lesivo verificatosi nella fase della programmazione tragga origine da un errore non riconoscibile. Si tratta dunque di casi in cui anche quei discutibili canoni individuati dalla giurisprudenza per l'imputazione del danno da prodotto, che giudicano sufficienti dei profili di tipicità oggettiva e soggettiva "affievoliti", risultano inadatti a reggere l'addebito dell'evento lesivo.

Questo secondo ordine di ipotesi assume un rilievo pratico di primaria importanza, ponendosi, in prospettiva, addirittura più significativo rispetto al primo e questo non solo per la frequenza statistica con cui un errore simile si potrebbe determinare all'interno

---

<sup>365</sup> C.Piergallini, *Danno da prodotto e responsabilità penale*, cit., pp. 99-100.

<sup>366</sup> *Ivi*, pp. 100 e ss.

<sup>367</sup> *Ivi*, p. 102.

dell'oscuro meccanismo di machine learning, ma prima ancora per i numerosi risvolti dilemmatici collegati all'irriconcoscibilità dell'errore che ha generato l'evento lesivo.

In simili scenari, insieme all'impossibilità di effettuare un qualsiasi giudizio di prevedibilità *ex ante*, sembrerebbe naufragare anche qualsivoglia prospettiva di imputare, anche solo per colpa, il fatto proprio del robot a un soggetto umano. Non solo, in base al principio dell'*oltre ogni ragionevole dubbio*<sup>368</sup>, anche tutti quei casi in cui non sia certa la natura del malfunzionamento dovrebbero necessariamente essere ricondotti alla più favorevole ipotesi della non imputabilità del fatto proprio del robot a soggetti umani<sup>369</sup>.

Sembrerebbe dunque residuare, nelle auto completamente autonome, un ventaglio di situazioni in cui il danno rimarrebbe riconducibile alla macchina stessa: casi che potrebbero irrimediabilmente essere destinati a rimanere penalmente irrilevanti<sup>370</sup>, prospettandosi dunque un inquietante vuoto di responsabilità<sup>371</sup>.

In questi casi, in cui la disciplina da tenere in considerazione è quella relativa alla colpa del programmatore per difetto di programmazione, due sono le strade che il legislatore potrebbe intraprendere.

- Da un lato, si potrebbe imputare, a chi immette nel mercato agenti intelligenti, ogni tipo di danno possibile, anche qualora sia prevedibile solamente in modo solo astratto e generico, ossia anche in assenza di leggi scientifiche consolidate e corroborate da studi, o addirittura nella totale assenza di ogni spiegazione causale. Dovrà di conseguenza ritenersi colposo il comportamento del programmatore *che non preveda l'imprevedibilità dell'agente artificiale!* L'astratta e generica prevedibilità di danni futuri, anche se non precisamente identificabili, incardinerebbe dunque un'ipotesi di responsabilità per colpa anche nel caso in cui non siano noti tutti gli anelli del processo causale. Il programmatore risponderà dunque a titolo di colpa per qualunque evento dannoso impre-

---

<sup>368</sup>Si tratta di un principio che permea l'intero ordinamento penal-processuale e che trova saliente espressione nelle garanzie fondamentali inerenti al processo penale quali la presunzione di innocenza dell'imputato, l'onere della prova a carico dell'accusa, l'enunciazione del principio *in dubio pro reo* e l'obbligo di motivazione e giustificazione razionale della decisione a norma degli artt. 111 c. 6 Cost. e 192 c. 1 c.p.p. (Cass. pen. sez. I 14 maggio 2004).

<sup>369</sup> A.Cappellini, *Profili penalistici delle self-driving cars*, cit., p. 340.

<sup>370</sup> S. Gless, E.Silverman, T. Weigend, *If robots cause harm, who is to blame? Self-driving cars and criminal liability*, cit., p. 432, che parla per l'appunto di "*responsibility gap*".

<sup>371</sup> A.Cappellini, *Profili penalistici delle self-driving cars*, cit., p. 340.

vedibile causato dall'agente artificiale purché costituisca il logico sviluppo delle modalità di funzionamento degli stessi agenti<sup>372</sup>. Resterebbe dunque una componente di rischio ineliminabile che ricadrebbe in capo alla società produttrice.

- Dall'altro lato si potrebbe ritenere che *l'uomo dietro la macchina* non sia mai responsabile per negligenza, non essendo sempre prevedibile l'agire di un robot che abbia la capacità di elaborare informazioni in modo indipendente ed autonomo, e quindi, di assumere conseguentemente decisioni altrettanto autonome<sup>373</sup>. Tale vuoto di responsabilità, o *responsability gap*<sup>374</sup>, si presenta come l'area più problematica da gestire sotto un profilo politico-criminale, nell'ottica dell'allocatione di un rischio, non naturale, ma tecnologicamente derivato e soprattutto generato dall'uomo stesso. Si tratta di un rischio destinato a ricadere, anzitutto, sulla vittima e sulla famiglia della stessa al pari di quanto avviene in caso di calamità naturali, o di eventi imputabili al caso o alla sorte avversa<sup>375</sup>. In realtà il rischio di cui trattasi ha natura ben diversa rispetto a quello relativo alle fatalità naturali, non prevedibili o non altrimenti impedibili da parte dell'uomo. Il rischio tecnologico infatti, seppur, come si è visto, non è prevedibile, è stato creato dall'uomo e sempre dall'uomo è stato tollerato: se circoleranno in futuro automobili self drivinig di livello 5 significa che è stato effettuato un giudizio di valore che stimi socialmente prevalente l'interesse a giovare dei benefici derivanti dall'uso della tecnologia driverless<sup>376</sup> rispetto agli eventuali danni, seppur esigui, che l'A.I. potrebbe causare, insieme all'interesse a individuare uno specifico soggetto responsabile.

Il suddetto bilanciamento permetterebbe dunque di salvare molte vite che in assenza dello sviluppo di tali tecnologie sarebbero andate perdute. L'altra faccia della medaglia è tuttavia rappresentata dal fatto che, in tale maniera, quelle vittime specifiche e concrete, anche se numericamente inferiori, ricadrebbero sulle spalle dell'intera società,

---

<sup>372</sup> Così M.B.Magro, A.I.: *la responsabilità penale per la progettazione, la costruzione e l'uso dei robot*, cit., p. 2.

<sup>373</sup> *Ibidem*.

<sup>374</sup> Questo il termine più volte utilizzato da S.Beck, *Google Cars, Software Agents, Autonomous Weapons Systems –New Challenges for Criminal Law*, in *Robotics, Autonomics and the Law*, a cura di E. Hilgendorf, U. Seidel, Nomos, Baden-Baden, 2017, pp. 227-251., nonché S. Gless, E. Silverman, T. Weigend, *If robots cause harm, who is to blame? Self-driving cars and criminal liability*, cit., pp. 412-436.

<sup>375</sup> A.Cappellini, *Profili penalistici delle self-driving cars*, cit., p. 340.

<sup>376</sup> Cfr. S.Gless, E.Silverman & T.Weigend, *If robots cause harm, who is to blame? Self-driving cars and criminal liability*. cit., pp. 430 ss.

una società che si è determinata nel senso di barattare la loro incolumità con quella della più ampia moltitudine che, in tal modo, verrebbe risparmiata<sup>377</sup>.

Si potrebbe quindi affermare che, una volta scomparsa la figura del conducente ad assorbire le colpe dei sinistri stradali, la tensione politica tra vantaggi sociali e istanze precauzionistiche si presenterà nella sua nudità. L'alternativa si presenta dunque secca: autorizzare e di conseguenza accettare che i beni in gioco possano in taluni casi rimanere sprovvisti di tutela o proibire radicalmente tale nuova tecnologia, perdendo così i benefici da questa apportati anche in termini di vite umane<sup>378</sup>.

Prevedere con precisione, in scenari tecnologici ma prima ancora sociali e politici tutto sommato ancora abbastanza lungi dal venire a esistenza, come la suddetta tensione verrà risolta rappresenta dunque una sfida particolarmente ardua.

Si è detto a tal proposito di come il tessuto sociale, e la percezione del rischio da parte dello stesso, possa giocare un ruolo di primo piano, anche se meno visibile, nello sviluppo della materia. Pertanto, si potrà immaginare come, almeno in un primo periodo, residuerà per ancora un certo lasso temporale la resistenza ad autorizzare veicoli completamente driverless nel nostro paese così come negli altri ordinamenti informati dai principi etici europei<sup>379</sup> e, qualora la circolazione delle stesse venga effettivamente autorizzata, si immagina che sia più probabile un indirizzo legislativo e giurisprudenziale orientato verso la prima delle due ipotesi esaminate. La circolazione di tali veicoli totalmente autonomi sarà tollerata, se non altro dal sentire sociale, a condizione che un eventuale sinistro non rimanga impunito. Si tratta comunque di una strada che sembrerebbe non tenere in considerazione uno dei principi cardine dell'intero diritto penale sancito altresì a livello costituzionale, il principio della personalità della responsabilità penale: il programmatore si

---

<sup>377</sup> A. Hevelke, e J. Nida-Rümelin, *Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis*, in «Science and Engineering Ethics», 21, 2015, p. 623, tuttavia, con riferimento al campo etico, osservano giustamente, che per qualunque misura di sicurezza ci possono essere dei casi concreti in cui essa faccia «più male che bene»; o addirittura sia causa di un danno, che in sua assenza non si sarebbe verificato. Si consideri, ad esempio, la cintura di sicurezza, per rimanere nell'ambito della circolazione stradale. Se di normale essa costituisce un presidio capace di ridurre i danni alle persone, in casi eccezionali può bloccarsi e imprigionare così i passeggeri, esponendoli senza possibilità di fuga alle fiamme di un incendio, o all'affogamento dentro il veicolo che affondi. Tuttavia, certamente «*questi tragici casi non cambiano il fatto che tali misure innalzino significativamente la sicurezza complessiva*» degli utenti della strada.

<sup>378</sup> A. Cappellini, *Profili penalistici delle self-driving cars*, cit., p. 340.

<sup>379</sup> Si rinvia a tal proposito alle profonde riflessioni riguardo l'«illusione prometeica» circa le possibilità di controllo del caso fortuito da parte della tecnica svolte da D. Castronuovo, *La colpa penale*, Giuffrè, Milano, 2009, pp. 92 ss.

troverebbe in definitiva nella condizione di dover rispondere per fatto altrui, ossia per il fatto proprio del robot<sup>380</sup>.

Nessun altro modo di proteggersi da tali rischi imprevedibili, ma ineliminabili, sembrerebbe altrimenti sussistere, se non attraverso l'opposta e radicale soluzione di proibire *ab origine* lo sviluppo di autonomous driving di livello 5 facendo appello a primarie istanze di precauzione<sup>381</sup>. In tema di produzione e distribuzione di prodotti potenzialmente pericolosi, l'ordinamento giuridico prevede infatti regole particolarmente severe, che sono state in gran parte trasferite dal diritto civile al diritto penale e che, in linea di principio, si ritengono applicabili anche ai prodotti dell'A.I., i cui rischi sono particolarmente elevati e difficili da controllare.

Per andare esente da responsabilità il produttore dovrebbe pertanto commercializzare solo quei prodotti il cui grado di sicurezza sia stato verificato e certificato prima della distribuzione dello stesso nel mercato.

Affinché dunque ci si determini nel senso di non immettere mai nel mercato le automobili di livello 5, il rischio in questione non dovrebbe essere annoverato tra i rischi tollerati, e pertanto, l'unica regola cautelare cui attenersi sarebbe rappresentata dall'obbligo di astenersi dall'esercizio di tale attività.

Tuttavia, la conseguenza di un atteggiamento eccessivamente cauto e prudente nella produzione e distribuzione di, almeno alcuni, agenti intelligenti a causa della loro mancanza di controllabilità in via precauzionale, si sostanzierebbe di fatto in un boicottaggio dell'uso di agenti intelligenti e di tutte le tecnologie più innovative<sup>382</sup>. Si rinunciarebbe pertanto a ogni beneficio derivante dallo sviluppo delle stesse di cui ne potrebbe beneficiare l'intera umanità<sup>383</sup>.

In una simile ottica di evoluzione e sviluppo a livello non solo economico ma anche sociale, appare dunque auspicabile che, almeno nel lungo periodo, anche nel quadro di un

---

<sup>380</sup> Del tema relativo alla possibile compressione dei principi fondamentali dell'ordinamento penale si è discusso nel cap 3 par 1.1.

<sup>381</sup> Per tutti, già L.Stortoni, *Angoscia tecnologica ed esorcismo penale*, in «Rivista Italiana di Diritto e Procedura Penale», 1, 2004, p. 83, nonché più specificamente S.Beck, *Google Cars, Software Agents, Autonomous Weapons Systems*, cit., p. 243.

<sup>382</sup> M.B.Magro, *A.I.: la responsabilità penale per la progettazione, la costruzione e l'uso dei robot*, cit., p.3.

<sup>383</sup> S. Gless et al., *If robots cause harm, who is to blame?*, cit., p. 430.

rigoroso rispetto dei principi fondamentali che reggono e circondano l'ordinamento penale, il legislatore possa orientarsi sulla seconda delle due ipotesi sopra analizzate e che prevalgano dunque istanze di tolleranza relative al rischio tecnologico, con la conseguente allocazione di tale ineliminabile componente di imprevedibilità all'interno dell'orbita del rischio consentito<sup>384</sup>.

Si tratta di una soluzione che ovviamente, come si è dimostrato poco sopra, non risulta priva di punti di criticità, ma che, se tale forma di tecnologia sarà destinata a entrare nella realtà giornaliera potrà costruire una fonte di rischio generalmente tollerata, alla stregua di un terremoto o di un'eruzione vulcanica.

Così come l'unica via che il legislatore può seguire al fine di proteggere la società da tali eventi naturali consiste ad esempio nella previsione di obblighi relativi a costruzioni antisismiche, allo stesso modo si potrà ipotizzare che egli provveda alla protezione da danni di machine learning operando su versanti diversi, primo tra tutti quello relativo alle infrastrutture.

Si consideri inoltre che, in realtà, una volta razionalizzata a livello antropologico e culturale l'esistenza del suddetto responsibility gap, l'esigenza di soddisfare le legittime pretese delle vittime potrebbe essere colmata su piani diversi rispetto a quello penalistico, primo fra tutti quello risarcitorio<sup>385</sup>, informato a schemi imputativi che, in ipotesi eccezionali, consentono di prescindere dai principi di responsabilità per fatto proprio e colpevolezza, principi che al contrario rappresentano la colonna portante della responsabilità penale.

---

<sup>384</sup> In questa direzione si muove anche S. Gless, E. Silverman e T. Weigend., *If robots cause harm, who is to blame?*, cit., pp. 433 ss.

<sup>385</sup> Secondo tradizionali meccanismi assicurativi di riallocazione dei risarcimenti sull'intera platea degli utenti della strada, che prima facie paiono meno problematici rispetto all'imputazione penalistica dei fatti, magari coadiuvati da fondi pubblici di risarcimento per le vittime da attivare nei casi problematici, da finanziare ricorrendo alla fiscalità generale.

### 3. RESPONSABILITÀ PER DANNI PROVOCATI DA SCELTE

#### 3.1 Introduzione

Come introdotto dal secondo capitolo di tale analisi, gli eventi dannosi cagionati dai prodotti dell'intelligenza artificiale possono essere ricondotti, oltre che ad errori di funzionamento del software, anche a scelte che la macchina stessa dovrà assumere in determinati casi. Si è sopra analizzato come in tale scenario la situazione tipica a venire in rilievo sia rappresentata dal cd. *dilemma del carrello*<sup>386</sup>, situazione in cui si richiede che gli algoritmi di machine learning, per il tramite del quale la macchina è in grado di agire, siano indirizzati dai programmatori ad assumere una decisione a discapito di un'altra.

È interessante notare come, intorno a tale dilemma, è stata sviluppata una branca molto particolare dell'etica analitica chiamata informalmente «*carrellologia*» (*trolleyology*)<sup>387</sup>, che si occupa per l'appunto di esperimenti mentali messi a punto allo scopo di indagare sulla capacità di prendere decisioni in condizioni estreme seppur immaginarie.

La difficoltà etica insita nella risoluzione di tale dilemma risiede nel fatto che, in simili circostanze, qualunque decisione presa dalla macchina condurrà inevitabilmente al sacrificio di vite umane o comunque di proprietà altrui.

L'obiettivo che in tale capitolo ci si prefigge di raggiungere è pertanto quello di inquadrare il problema della responsabilità penale nel caso in cui l'evento da cui scaturisce il danno sia provocato da una scelta assunta in sede di programmazione che, in situazioni quali quelle prospettate dal *trolley problem*, imponga necessariamente il sacrificio di una o più vite al fine di salvarne altre.

Più in particolare, nel perseguire tale obiettivo, si dovrà tener presente la fondamentale circostanza secondo la quale la qualifica di prodotto autonomo modifica sostanzialmente i termini entro cui si dovrà inquadrare il problema della responsabilità in simili situazioni. Nelle automobili tradizionali, controllate esclusivamente dall'azione umana, tale scelta

---

<sup>386</sup>Le formulazioni originali dell'apologo del tram P. Foot, *The Problem of Abortion and the Doctrine of the Double Effect in Virtues and Vices and Other Essay in Moral Philosophy*, cit. e di J. JarvisThomson, *The Trolley Problem*, in «The Yale Law Journal», 5, 1985, pp. 1395-1415.

<sup>387</sup> Il termine fu coniato dal filosofo britannico Kwame Anthony Appiah. Si veda a tal proposito: K. A. Appiah, *Experiments in Ethics*, Harvard University Press, Cambridge (MA), 2010.

verrà compiuta direttamente dall'automobilista in una situazione di pericolo estremo e soprattutto attuale. Ciò comporta che, in sede di giudizio penale, potrebbe essere applicata la scriminante dello *stato di necessità*<sup>388</sup>. Al contrario, in un'automobile a guida autonoma, la scelta che la stessa implementerà sarà una decisione standardizzata e pre-programmata, circostanza questa che induce ad escludere la possibilità di applicare la detta causa di esclusione della responsabilità, comportando così problematiche sul piano giuridico, oltre che sicuramente sul piano etico. Ogni tipo di scelta consapevole infatti, e a maggior ragione quelle che devono essere assunte in situazioni come quella sopra prospettata, presuppongono un'etica<sup>389</sup>, la cui individuazione appare un importante punto di riferimento per affrontare il problema della responsabilità ed eventuali cause di esclusione della stessa.

Al fine di procedere correttamente nell'obbiettivo di inquadrare, in simili situazioni, il problema della responsabilità penale, si propone pertanto di partire dall'analisi di un'eventuale allocazione della responsabilità nel caso in cui tipi di veicoli diversi siano coinvolti in incidenti inevitabili, considerando così le differenze esistenti sul piano giuridico nel caso in cui venga in rilievo un'automobile tradizionale e un'automobile a guida autonoma.

La successiva e più specifica analisi circa il problema dell'allocazione della responsabilità penale per queste ultime verrà poi condotta adottando un approccio giuridico-filosofico, rivolgendo uno sguardo critico a come la legge abbia già regolato scelte altrettanto difficili in simili scenari di emergenza. In particolare, si propone di considerare la dottrina giuridica sviluppatasi intorno allo stato di necessità quale promettente punto di partenza: essa è infatti rivolta a regolare quei casi di emergenza in cui evitare tutti i mali è ritenuto impossibile e in cui agenti umani hanno causato intenzionalmente danni alla vita e alla proprietà altrui al fine di evitare altri danni e perdite. Partendo da una ricostruzione di

---

<sup>388</sup> La scriminante è prevista dall'art. 54 c.p. e risulta applicabile anche in caso di commissione di reati colposi. Per completezza si riporta di seguito il testo completo dell'articolo: "*Non è punibile chi ha commesso il fatto per esservi stato costretto dalla necessità di salvare sé od altri dal pericolo attuale di un danno grave alla persona, pericolo da lui non volontariamente causato, né altrimenti evitabile, sempre che il fatto sia proporzionato al pericolo. Questa disposizione non si applica a chi ha un particolare dovere giuridico di esporsi al pericolo. La disposizione della prima parte di questo articolo si applica anche se lo stato di necessità è determinato dall'altrui minaccia; ma, in tal caso, del fatto commesso dalla persona minacciata risponde chi l'ha costretta a commetterlo*".

<sup>389</sup> Così J. F. Bonnefon, A. Shariff, I. Rahwan, *The Social Dilemma Of Driverless Cars*, cit. pp. 1573.

alcuni principi e norme fondamentali incorporati nella dottrina e nella giurisprudenza anglo-americana sulla "*dottrina della necessità*", si potrà così valutare, se del caso, quali di questi possano effettivamente essere utilizzati per trovare linee guida ragionevoli al fine di risolvere la questione etica e giuridica della regolamentazione della programmazione dei veicoli autonomi in situazioni di emergenza.

Come più volte riferito nel corso della trattazione, ogni tipo di scelta, soprattutto se assunte in situazioni estremamente complicate, presuppongono un'etica; così, come risulta difficile risolvere il dilemma del carrello da un punto di vista puramente morale ed etico, risulterà ancora più complesso risolverlo dal punto di vista giuridico. Si tratta in realtà di un circolo vizioso: la questione giuridica non può essere risolta se prima non si è risolta la questione etica. Pertanto, all'interno di questa analisi, anziché considerare separatamente le due questioni si propone di utilizzare gli elementi sia etici che legali al fine di analizzarle e tentare di risolverle congiuntamente.

Secondo il filosofo John L. Austin<sup>390</sup>, il ragionamento giuridico può infatti costituire un efficace strumento di chiarimento e un nuovo punto di partenza per risolvere complicate questioni filosofiche stante la sua necessità di apprestare una costante attenzione alle situazioni che si prospettano nella vita reale, di raggiungere risposte chiare e ben fondate e, infine di fondare una riflessione che sia in parte indipendente da teorie e propensioni astratte tipiche della filosofia pura. La riflessione filosofica può quindi trarre vantaggio dal considerare principi e norme legali in quanto questi ultimi rappresentano in molti casi il risultato di una combinazione tra principi morali astratti e considerazioni pratiche derivanti dall'osservazione e dall'analisi comparativa di casi reali; inoltre, le norme giuridiche costituiscono spesso un "*ragionevole compromesso tra principi e interessi in contrasto*"<sup>391</sup>

---

<sup>390</sup> J.L. Austin, *Una richiesta di scuse*, in *Philosophical Papers*, a cura di J. O. Urmson e G. J. Warnock, Clarendon Press, Oxford, 1961.

<sup>391</sup> H. Hart, *Il concetto di legge*. Clarendon Press, Oxford, 1961. Un approccio simile viene adottato anche da P. Asaro, *Un corpo da calciare, ma ancora nessuna anima da dannare: prospettive legali sulla robotica*, in *Robot Ethics* a cura di P. Lin, K. Abney, G. Bekey, MIT Press, Cambridge-Massachusetts, 2012, pp. 169-186.

### 3.2 Analisi di come varia la allocazione della responsabilità in scenari di emergenza a seconda del grado di autonomia dell'agente artificiale

Si illustrerà in questo paragrafo come la legge attualmente regola il comportamento dei conducenti umani nei dilemmi tra la vita e la morte e si tenterà quindi di ricostruire come questa, in un futuro più o meno prossimo, regolerà la responsabilità qualora tale dilemma sia stato risolto in sede di progettazione di software eticamente pre-programmati.

I dilemmi etici e legali sollevati in scenari riconducibili al cd. *Trolley problem* che seguiranno nei successivi paragrafi verranno trattati e affrontati prendendo in considerazione tre scenari illustrati da Bonnefon inevitabilmente implicanti un danno inevitabile imminente<sup>392</sup>:

1. rimanere in rotta e uccidere diversi pedoni o sterzare e uccidere un passante.
2. rimanere sulla rotta e uccidere un pedone o sterzare e uccidere il proprio passeggero.
3. rimanere in rotta e uccidere diversi pedoni o sterzare e uccidere il proprio passeggero.

#### IL PROBLEMA DELLA RESPONSABILITÀ NELLE AUTOMOBILI TRADIZIONALI

Supponiamo innanzitutto che l'auto sia guidata da un essere umano e che esso non abbia contribuito a creare lo stato di pericolo. In tale situazione, i tre scenari sopra descritti da Bonnefon, possono essere tutti legalmente giustificati in modo tale che il conducente umano possa evitare di incorrere in responsabilità penale.

Nella prima situazione, la scelta di rimanere in rotta e di far uccidere diversi pedoni, piuttosto che sterzare e uccidere un passante, potrebbe essere giustificata adottando una posizione morale-legale che condanni la causalità intenzionale della morte: l'automobilista in tale maniera non provocherebbe intenzionalmente la morte di nessuno<sup>393</sup>. Tale considerazione ovviamente risulta valida sempre che si assuma che, nonostante la situazione di estremo pericolo, il conducente riesca a mantenere la lucidità della scelta.

---

<sup>392</sup> J.F. Bonnefon, A. Shariff, I. Rahwan, *Autonomous vehicles need experimental ethics: are we ready for utilitarian cars?*, cit. pp. 1573 e ss.

<sup>393</sup> G. Contissa, F. Lagioia, e G. Sartor, *The Ethical Knob: ethically-customisable automated vehicles and the law*, cit., pp. 366 e ss.

Nel secondo scenario, la scelta di rimanere in rotta potrebbe invece essere giustificata invocando lo stato di necessità in quanto, non solo la situazione di pericolo si presenta come attuale, ma anche in quanto tale scelta si presenta come necessaria al fine di salvare la vita del conducente stesso. La causa di esclusione della responsabilità basata sullo stato di necessità è regolata in maniera diversa nelle diverse giurisdizioni, tuttavia, due requisiti risultano comuni:

- l'esistenza di un pericolo attuale, o quanto meno imminente, di gravi danni fisici al soggetto non causati volontariamente dallo stesso e comunque non altrimenti evitabili.
- la proporzionalità tra il fatto commesso e il pericolo affrontato.

Entrambi requisiti che possono ritenersi presenti nella situazione sopra descritta.

Lo stesso ragionamento potrebbe affermarsi anche prendendo in considerazione il terzo scenario, ossia qualora la scelta del conducente di salvare la propria vita condurrebbe alla morte di diverse persone<sup>394</sup>.

#### IL PROBLEMA DELLA RESPONSABILITÀ NEL CASO DI A.V. PRE-IMPOSTATI ETICAMENTE

Nel momento in cui vengono in rilievo le automobili driverless non si può prescindere dall'effettuare considerazioni diversificate tra i diversi livelli di autonomia individuati dalla SAE International: la presenza a bordo di un conducente umano, sopra definito come *conducente potenziale*, sembrerebbe richiedere l'applicabilità delle stesse considerazioni svolte nel paragrafo precedente con riferimento alle automobili tradizionali. Al contrario, il veicolo totalmente autonomo di livello 5, benché allo stato non ancora sviluppato, richiederà invece valutazioni diametralmente opposte stante l'assenza di un soggetto umano chiamato ad assumere una scelta in una situazione di pericolo attuale e stante la programmazione *ex ante* del comportamento che assumerà il veicolo.

Supponendo infatti che il comportamento della macchina sia stato programmato *ex ante* in sede di progettazione e programmazione del software, le considerazioni sopra svolte, relative all'assenza di responsabilità o di punibilità<sup>395</sup> del conducente umano, non sembrerebbero potersi applicare nell'ipotesi in questione. Si tratta infatti di una situazione in

---

<sup>394</sup> *Ibidem*.

<sup>395</sup> Si è ritenuto di inserire entrambi i termini "responsabilità" e "punibilità" in quanto la natura dello stato di necessità, come si analizzerà ampiamente in seguito, non è pacifica in dottrina. Nel caso in cui si assuma che esso sia una scusante si assisterà ad una assenza di colpevolezza, nel caso in cui si consideri lo stato di necessità quale causa di giustificazione verrebbe meno l'esigenza punitiva dello stato.

cui il pericolo, seppur inevitabile, non si presenta come attuale e tale circostanza rende difficoltosa l'applicazione degli istituti penalistici sopra esaminati.

Nello specifico, se nel primo dei tre scenari ipotizzati da Bonnefon il guidatore umano può risultare legalmente giustificato qualora scelga di rimanere in rotta, sacrificando così la vita di più pedoni piuttosto che sterzare e uccidere un solo passante, tale soluzione non appare applicabile qualora tale scelta sia assunta, *ex ante*, in sede di programmazione.

In effetti, la distinzione tra omettere di intervenire, lasciando così che l'automobile prosegua nel suo percorso, e agire in un determinato modo, scegliendo al contrario di sterzare - una distinzione che, come analizzato poco sopra nel caso di un'auto con equipaggio, può giustificare la scelta umana di consentire all'auto di mantenere la rotta seguita - non sembrerebbe potersi applicare al programmatore, in quanto quest'ultimo sceglierebbe deliberatamente di sacrificare un numero maggiore di vite. Quella posizione morale-legale tesa a condannare la causalità intenzionale della morte non sembrerebbe pertanto poter essere adottata, in tale scenario, dal programmatore<sup>396</sup>.

Le problematiche esaminate appaiono persistere anche nell'analisi del secondo scenario. Il produttore o il programmatore in tal caso, non affronta direttamente un pericolo imminente per la sua vita, ma piuttosto adotta preventivamente una specifica scelta allo scopo di salvare una o più persone, provocando al contempo, inevitabilmente, danni ad altri soggetti coinvolti nella medesima situazione pericolosa. L'applicabilità dello stato di necessità è infatti controversa qualora l'autore non sia direttamente in pericolo e non agisca per autoconservazione o conservazione di altre persone a lui vicine<sup>397</sup>.

La responsabilità del programmatore per le scelte adottate dall'algoritmo di machine learning dipenderà dunque da una futura estensione giurisprudenziale e dottrinale della scusante in questione. Se la giurisdizione legale consentisse in tale scenario l'applicabilità dello stato di necessità, il programmatore non sarebbe infatti punibile per nessuna delle due scelte, e ciò in base agli stessi motivi esaminati con riferimento alle ipotesi di responsabilità relative all'automobilista umano. In caso contrario, non sarebbe legalmente accettabile una pre-programmazione della macchina sia in una direzione che nell'altra: in

---

<sup>396</sup> G. Contissa, F. Lagioia, G. Sartor, *The Ethical Knob*, cit., p. 368.

<sup>397</sup> *Ivi*, p. 368 e ss.

entrambi i casi si porrebbe in capo al programmatore una scelta ingiustificata ed arbitraria tra due vite<sup>398</sup>.

Santoni de Sio, sul quale si ritornerà in seguito, rifiuta qualsiasi pre-programmazione in senso utilitaristico di un AV, ossia, in tal caso, l'adozione ex ante della scelta di sterzare volontariamente al fine di salvare il maggior numero di vite. Egli fa leva sul dato secondo cui la legge generalmente non consente di uccidere una persona innocente al fine di salvare la vita di altre persone<sup>399</sup>.

Il terzo scenario infine, amplia le problematiche esaminate con riferimento alla situazione preesistente: la pre-programmazione dell'auto nel senso di continuare la sua traiettoria, causando la morte di un numero più elevato di persone, appare sia moralmente che legalmente non giustificata e non giustificabile in nessuna giurisdizione; equivarrebbe infatti a una scelta arbitraria per uccidere molti piuttosto che uno<sup>400</sup>.

In definitiva l'analisi dei tre scenari ipotizzati da Bonnefon evidenzia come determinate scelte pre-programmate risulterebbero moralmente e legalmente poco accettabili, anche qualora le corrispondenti scelte assunte dal conducente umano verrebbero al contrario approvate, o quantomeno scusate, da un punto di vista legale.

### **3.3 Stato di necessità e dibattito sull'operatività dello stesso con riferimento alla pre-programmazione degli AV**

Sebbene a primo sguardo la risoluzione ex ante, in sede di progettazione, di un difficile dilemma etico quale il *trolley problem* potrebbe apparire alla stregua di un circolo vizioso, si noti come la legge, in realtà, abbia già regolato scelte altrettanto difficili assunte in simili scenari di emergenza.

Si propone dunque, come accennato nel paragrafo introduttivo, di considerare la dottrina giuridica della necessità quale promettente base di ragionamento per l'analisi in questione: tale dottrina rivolge infatti l'attenzione a quei casi in cui agenti umani hanno causato intenzionalmente danni alla vita e alla proprietà altrui al fine di evitare altri danni e

---

<sup>398</sup> *Ibidem*.

<sup>399</sup> F. Santoni de Sio, *Killing by Autonomous Vehicles and the Legal Doctrine of Necessity*. in «Ethic Theory Moral Practice», 2, 2017, pp. 415 e ss.

<sup>400</sup> G. Contissa, F. Lagioia, G. Sartor, *The Ethical Knob*, cit. 370 e ss.

perdite in situazioni di emergenza in cui evitare tutti i mali risulterebbe comunque impossibile. Lo scopo sarà dunque quello di valutare se, e del caso quali, di questi principi e norme possano effettivamente essere utilizzati al fine di ricercare linee guida ragionevoli per la risoluzione della questione della regolamentazione della responsabilità in caso di pre-programmazione dei veicoli autonomi in simili situazioni di emergenza.

Si ritiene doveroso premettere che, al fine di applicare qualsiasi difesa, è necessario che in prima facie sia stato commesso un reato; e, come analizzato nel par x, un comportamento può essere valutato alla stregua di un'offesa penalmente rilevante solamente nel caso in cui risultino sussistenti i due elementi essenziali del reato: l'actus reus e la mens rea.

Si è analizzato sopra come sia attualmente controversa la possibilità di attribuire la responsabilità penale per i danni provocati dai prodotti dell'intelligenza artificiale, se direttamente al prodotto stesso o al contrario all'agente umano dietro la macchina; in tale paragrafo l'attenzione verrà focalizzata esclusivamente su scenari in cui possa affermarsi in modo sicuro che il programmatore abbia intenzionalmente impostato il comportamento di un veicolo in una situazione di emergenza, sulla base della ragionevole convinzione che il veicolo si comporterà di conseguenza. Si tratta quindi di scenari in cui gli agenti umani dietro le macchine possano legittimamente essere ritenuti responsabili per il comportamento del veicolo automatizzato e che, di conseguenza, possano essere chiamati a rispondere per i danni o le perdite causati dal comportamento intelligente e pre-programmato di un veicolo<sup>401</sup>.

Ritornando alla cosiddetta *dottrina della necessità*, essa riguarda le legittime eccezioni relative al rispetto di una norma proibitiva legale. Così, comportamenti in prima facie vietati dal diritto penale, possono essere ammessi e giustificati in situazioni eccezionali. La necessità si presenta dunque come una difesa: un'affermazione che un imputato può sollevare per dimostrare che il suo comportamento, sebbene generalmente proibito dalla legge penale, non dovrebbe essere considerato un crimine sulla base di tutte le circostanze presenti nel caso concreto<sup>402</sup>.

---

<sup>401</sup> Se tale ipotesi sulla *tracciabilità* della responsabilità sia ragionevole in relazione a tutti i possibili scenari futuri che coinvolgono agenti autonomi è una questione controversa, si veda ad esempio il cosiddetto problema del "responsability gap" individuato dal Parlamento Europeo nella risoluzione del 17 febbraio 2017.

<sup>402</sup> Cfr. M.S. Moore, *Atto e criminalità: la filosofia di azione e le sue implicazioni per il diritto penale*, Oxford University Press, New York, 2010.

Nel codice penale italiano lo stato di necessità viene regolato dall'articolo 54, il quale indica come presupposto necessario e imprescindibile “*una costrizione dalla necessità di salvare se od altri dal pericolo attuale di un danno grave alla persona*”. Sennonché, come ripetuto più volte nel corso della trattazione di tale capitolo, nel caso di programmazione di un veicolo autonomo l'elemento dell'attualità del pericolo risulta essere assente e ciò pone problemi di compatibilità con l'applicazione dell'art. 54 nell'ipotesi di pre-programmazione del comportamento degli AV in scenari di emergenza.

Tuttavia, si ritiene necessario rilevare come l'operatività dello stato di necessità nell'ipotesi presa in considerazione potrebbe in larga parte risultare influenzata dallo stesso inquadramento giuridico dell'articolo 54 c.p.: se una larga corrente dottrina considera lo stato di necessità alla stregua di una scusante, un'opposta teoria interpretativa tende, al contrario, ad inquadrarlo tra le cause di giustificazione<sup>403</sup>. Il soggetto che pone in essere l'azione penalmente rilevante deve infatti essere *costrretto* dalla necessità di commettere quel fatto e, a tal proposito, ai fini dell'inquadramento dell'articolo 54 c.p., giocherà un ruolo primario il significato da attribuire al termine “costrizione”<sup>404</sup>.

#### STATO DI NECESSITÀ COME SCUSANTE

Analizzando distintamente le due diverse interpretazioni, si noti come una larga corrente dottrina ritiene che lo stato di necessità dovrebbe essere inquadrato alla stregua di una scusante<sup>405</sup>, vale a dire una causa che esclude la colpevolezza nei confronti di chi abbia agito sotto l'effetto di un'enorme pressione psicologica la quale, agli occhi del legislatore, rendeva inesigibile un diverso comportamento rispettoso dei precetti penali<sup>406</sup>. Una simile lettura consentirebbe di spiegare il motivo per cui il suo ambito di applicazione è in genere ristretto a scenari straordinari e improvvisi, in cui gli agenti sono in preda al panico o

---

<sup>403</sup>Vi è in realtà una terza corrente dottrina che considera lo stato di necessità in un duplice modo distinguendo l'ipotesi in cui venga sacrificato un bene dal valore sociale minore di quello minacciato e il caso in cui venga offeso un bene di pari valore. Nel primo dei due casi, secondo tale interpretazione, mancherebbe l'antigiuridicità del fatto. Nel secondo sarebbe assente l'elemento della colpevolezza e pertanto sarebbe inquadrato alla stregua di una scusante.

<sup>404</sup>G. Marinucci, E. Dolcini, G. Gatta, *Manuale di diritto penale. Parte generale*, VIII ed., Giuffrè, Milano, 2018, cit. p. 311.

<sup>405</sup>Questa è l'opinione da tempo dominante nella dottrina tedesca, patrocinata inizialmente da J. P. Goldschmidt, *Der Notstand, ein Schuldproblem*, Manzsche k.u.k. Hof-Verlags- und Universitäts-Buchhandlung, 1913. così come citato da E. Mezger, *Diritto Penale*, cit., p. 248. Per la dottrina italiana si vedano L. Scarano, *La non esigibilità nel diritto penale*, Casa Editorial Libreria Humus, Napoli, 1948, pp. 126 e ss. e G.D. Pisapia, *Istituzioni di diritto penale*, CEDAM, Padova 1965 p. 180.

<sup>406</sup>G. Marinucci, E. Dolcini, G. Gatta, *Manuale di diritto penale. Parte generale*, cit., p. 311.

comunque sottoposti a un'eccezionale pressione psicologica determinata dalla situazione di pericolo: gli agenti che operano *in necessità* non sono colpevoli in quanto la loro volontà è sospesa e quindi le loro azioni devono essere considerate moralmente involontarie<sup>407</sup>.

Secondo tale interpretazione infatti, la *costrizione* richiesta dall'articolo 54 viene identificata con l'esclusione, o quantomeno con una forte limitazione della capacità di agire, presupponendo a tal proposito, in capo a colui che pone in essere l'azione, non solo la consapevolezza del pericolo, ma altresì un effettivo turbamento psicologico<sup>408</sup>.

A sostegno dell'inquadramento dello stato di necessità in particolare militano diverse argomentazioni:

- in primo luogo si prenda in considerazione un dato di carattere empirico: tutti i casi tradizionalmente ricondotti sotto lo stato di necessità sono caratterizzati da un'effettiva pressione psicologica determinata da eventi esterni, chiamando addirittura in causa *l'istinto di conservazione*.
- si consideri inoltre che, solo tramite una simile lettura del termine "costrizione" inserito nell'art. 54 c.p., una lettura quindi che dia risalto al turbamento motivazionale dell'agente, si eviterebbe di ricondurre sotto la protezione dello stato di necessità una serie di casi universalmente considerati meritevoli di sanzione penale, come ad esempio il caso del medico che, disponendo di una sola apparecchiatura di rianimazione, la destina al paziente più giovane giunto in ospedale successivamente togliendola al paziente più anziano che già ne usufruiva.
- la terza argomentazione si ricava infine dal tenore del terzo comma dell'articolo 54 c.p. relativo allo "*stato di necessità determinato dall'altrui minaccia*". Si tratta questa di un'ipotesi in cui la natura di scusante è pacifica: il soggetto che pone in essere l'azione penalmente rilevante è infatti vittima di una minaccia che limita la sua capacità di scelta e ne sospende la volontà. Stante l'inequivocabile tenore letterale che lega il primo e il terzo comma<sup>409</sup> la differenza tra i due risiede esclusivamente nella fonte del pericolo, che nel primo comma è rappresentata da un evento naturale o una condotta umana, mentre nel terzo comma esclusivamente da una condotta umana che assuma le vesti di

---

<sup>407</sup> G. P. Fletcher, *Rethinking criminal law*, Oxford University Press, New York, 2000.

<sup>408</sup> G. Marinucci, E. Dolcini, G. Gatta, *Manuale di diritto penale. Parte generale*, cit., p. 311.

<sup>409</sup> L'art 54 c.p. recita infatti che "*la disposizione della prima parte di questo articolo si applica anche se lo stato di necessità è determinato dall'altrui minaccia*".

una minaccia. Se pertanto è pacifico che il terzo comma dell'art. 54 c.p. sia una scusante, altrettanto dovrebbe esserlo per l'ipotesi disciplinata dal primo comma.

In definitiva il giudice dovrà sempre accertare che il soggetto agente abbia subito un *effettivo turbamento motivazionale*.

Appare importante notare come tale ricostruzione si riflette anche sull'ipotesi, che qui interessa, di "soccorso di necessità" ossia quei casi in cui l'agente ponga in essere l'azione al fine di salvare *altri* dal *pericolo attuale di danno grave alla persona*. La scusante in parola potrebbe quindi essere applicata solamente qualora la rappresentazione del pericolo incombente sul terzo abbia provocato un effettivo turbamento emozionale sul processo decisionale dell'agente. Per questo motivo si ritiene applicabile solamente qualora "l'altro" sia un soggetto vicino a colui che pone in essere l'azione penalmente rilevante, quale un parente, un coniuge, un convivente, un amico fraterno e così via<sup>410</sup>.

Se quindi si dovesse intendere la necessità in tal senso, allora essa non potrebbe probabilmente applicarsi nel caso in cui la decisione dannosa sia stata assunta in sede di programmazione di un AV: sarebbe sempre inammissibile istruire deliberatamente, in anticipo, un agente artificiale allo scopo di danneggiare qualcuno. Il programmatore infatti, lontano dalla situazione di pericolo, non risulta sottoposto a un effettivo turbamento psicologico determinato da eventi esterni, ne tantomeno potrebbe esserlo stante il fatto che, in sede di programmazione egli si troverebbe a decidere sulla vita e la morte di soggetti non solo sconosciuti ma soprattutto non identificabili a priori, soggetti quindi a lui non legati da alcun vincolo affettivo.

#### **STATO DI NECESSITÀ COME CAUSA DI GIUSTIFICAZIONE**

Qualora, al contrario, si inquadrasse l'art 54 c.p. tra le cause di giustificazione il discorso potrebbe in parte cambiare.

Le cause oggettive di esclusione del reato, comunemente dette cause di giustificazione rappresentano situazioni nelle quali un fatto, di regola vietato dall'ordinamento, viene consentito o addirittura imposto dall'ordinamento, risultando di conseguenza esente da pena in quanto non più considerato antiggiuridico. La ratio di tali cause è ravvisabile pertanto nella *mancaza di danno sociale*: l'azione non contrasta, alla ricorrenza di tali

---

<sup>410</sup> G.Marinucci, E. Dolcini, G.Gatta, *Manuale di Diritto Penale, Parte Generale*, cit., p. 312.

cause, con gli interessi della comunità, in quanto si presenta come necessaria al fine di salvaguardare un bene dal valore sociale maggiore o perlomeno uguale a quello sacrificato. Esulando il danno sociale, l'intervento dello stato tramite la sanzione punitiva non avrebbe più ragion d'essere<sup>411</sup>.

Se dunque si considerasse lo stato di necessità come causa di giustificazione verrebbe meno la rilevanza della pressione psicologica subita dal soggetto agente -pressione che senza dubbio risulta assente in capo al programmatore- ma al contrario, acquisirebbe rilievo l'impossibilità di salvare il bene esposto in pericolo senza dover sacrificare il bene di un terzo innocente. In altre parole, ciò che verrebbe in rilievo è rappresentato esclusivamente da un bilanciamento di beni in conflitto<sup>412</sup>, bilanciamento che, in via ipotetica, potrebbe essere richiesto al programmatore qualora sia chiamato a decidere su situazioni riconducibili al c.d. *dilemma del carrello*.

Tuttavia, sembrerebbe che, anche inquadrando lo stato di necessità tra le cause di giustificazione, l'art. 54 c.p. non possa comunque essere applicato alle ipotesi di danni provocati da scelte etiche assunte in sede di programmazione: non solo risulta in tale fase essere assente l'elemento dell'*attualità del pericolo* ma, prima ancora, le linee guida etiche per le self driving cars stilate dal ministero dei Trasporti tedesco, le più dettagliate esistenti in materia, sanciscono il principio per cui "*Non è possibile decidere ogni volta, e soprattutto a priori, se salvare una vita piuttosto che un'altra*"<sup>413</sup>. Si evince dunque che un bilanciamento tra beni in conflitto non è ammissibile qualora tale bilanciamento sia effettuato ex ante, in assenza di un pericolo attuale che minaccia l'integrità di un bene in pericolo.

#### **STATO DI NECESSITÀ IN ASSENZA DELL'ATTUALITÀ DEL PERICOLO**

In realtà la legge e la giurisprudenza consentono già alcuni danni intenzionali, anche alcuni omicidi, dopo misurati esercizi di giudizio e quindi in assenza dell'attualità del pericolo<sup>414</sup>. Per analizzare tali ipotesi è però necessario rivolgere l'attenzione oltre i confini

---

<sup>411</sup> F. Antolisei, *Manuale di Diritto Penale, Parte Generale*, cit., pp. 272-273.

<sup>412</sup> G. Marinucci, E. Dolcini, G. Gatta, *Manuale di Diritto Penale, Parte Generale*, cit., p. 311.

<sup>413</sup> Federal Ministry of Transport and Digital Infrastructure, *Ethic commission, Automated ad Connected Driving*, Report, Giugno 2017.

<sup>414</sup> Santoni de Sio, F. *Killing by Autonomous Vehicles and the Legal Doctrine of Necessity*. cit. pp. 415 e ss.

della giurisprudenza nazionale concentrandosi, in maniera particolare sulla giurisprudenza anglo-americana: in una società sempre più globalizzata e interconnessa infatti non risulta possibile prescindere da un'analisi comparatistica tra i diversi ordinamenti.

In particolare, nell'ordinamento inglese, ove da sempre la legge ha adottato un approccio restrittivo all'interpretazione della dottrina della necessità, ad oggi, come dimostrato da recenti casi, sta si sta mostrando un'apertura in tal senso. Così, se nel 1884, nel caso *R v Dudley e Stephens*, i marinai Tom Dudley e Edwin Stephens furono condannati per omicidio, nel 2001, con il *Re A* (gemelli siamesi) la situazione è iniziata a mutare.

In particolare, nel primo dei due casi, due marinai, naufragati insieme ad altri uomini e pensando che sarebbero presto morti di fame, uccisero uno degli altri compagni che versava in terribili condizioni di salute allo scopo di cibarsene. Nella condanna, il tribunale non tenne in considerazione né valutazioni di carattere utilitaristico, né escluse la responsabilità degli stessi nonostante la loro azione fosse stata posta in essere in uno stato di necessità. La condanna fu motivata dal principio per cui in nessuna circostanza nessuna vita innocente dovrebbe essere sacrificata a discapito di altre<sup>415</sup>.

Dopo quasi 120 anni, nel 2001, con il *Re A* (gemelli siamesi)<sup>416</sup>, la Corte adottò un approccio diametralmente opposto rispetto a quello seguito nella vicenda sopra analizzata. Il caso vedeva coinvolti Gracie e Rosie Attard, gemelli siamesi uniti all'addome. Le prove mediche indicavano che Gracie era il fratello più forte che stava sostenendo la vita di Rosie e che, qualora i gemelli fossero stati separati chirurgicamente, Gracie avrebbe avuto un tasso di sopravvivenza del 94% mentre Rosie non sarebbe sopravvissuta. Se tuttavia fossero rimasti congiunti, Gracie sarebbe dovuta morire prima di compiere sei mesi di vita, causando di conseguenza anche la morte di Rosie.

La Court of Appeal autorizzò in tal caso l'intervento chirurgico e, a sostegno di tale decisione, uno dei giudici, Lord Justice Brooke, invocò esplicitamente la necessità come difesa al fatto di aver consentito un intervento volto inevitabilmente a determinare la morte di uno dei gemelli allo scopo di salvare l'altro. Per quanto innovativo potesse essere stato per la legge inglese, *Re A* consentiva pertanto un'uccisione intenzionale per necessità in circostanze molto specifiche, vale a dire:

---

<sup>415</sup> Cfr. gli atti del processo, *Queen vs. Dudley and Stephens*, 14 Queens Bench Division 273, 9 dicembre 1884. Per un approfondimento sulla questione, e le conseguenti implicazioni di carattere filosofico, in particolare sul dibattito tra dottrine utilitaristiche e deontologiche, si veda M.Sandel, *Giustizia, il nostro bene comune*, cit., pp. 40-43.

<sup>416</sup> *Re A* (conjoined twins) (2001) 2 WLR 480

- la persona uccisa impediva, seppur involontariamente, la sopravvivenza dell'altra.
- la persona uccisa sarebbe sicuramente morta in breve tempo.
- l'omicidio è stato commesso con l'autorizzazione ufficiale di un'autorità pubblica (cioè un giudice), dopo un'attenta valutazione dei fatti pertinenti e il parere di esperti<sup>417</sup>.

Ci si chiede dunque, tralasciando l'ultima delle tre condizioni in base alle quali la giurisprudenza inglese ha ritenuto consentita l'uccisione intenzionale, se sia possibile immaginare, almeno in linea di principio, scenari di emergenza, che vedono coinvolti dei veicoli a guida autonoma, con caratteristiche analoghe a *Re A*.

Nel caso che vide coinvolti i due gemelli siamesi infatti risultava essere assente l'elemento dell'attualità del pericolo. Il pericolo infatti, seppur imminente, non potrebbe essere considerato attuale nonostante, in assenza dell'intervento, uno dei due gemelli sarebbe dovuto morire in ogni caso, comportando conseguentemente la morte dell'altro gemello entro un periodo di tempo limitato<sup>418</sup>.

È sulla scorta della distinzione tra *attualità* e *imminenza* del pericolo che il codice penale francese, nell'articolo 122-7, disciplina lo stato di necessità. La scriminante in esame presenta dunque un ambito di operatività decisamente più ampio rispetto quello tracciato dall'articolo 54 del codice penale italiano: il pericolo non deve presentarsi necessariamente come attuale, potendo essere anche solo imminente<sup>419</sup>.

Tuttavia, anche ipotizzando un'estensione dell'articolo 54 nella stessa direzione intrapresa dal codice penale francese, lo stato di necessità non potrebbe ritenersi applicabile al caso di pre-programmazione etica degli AV: il pericolo in questi casi non può considerarsi nemmeno imminente, ma si presenterebbe, al più, come meramente ipotetico.

---

<sup>417</sup> Santoni de Sio, F. *Killing by Autonomous Vehicles and the Legal Doctrine of Necessity*, cit. pp. 415 e ss.

<sup>418</sup> *Ivi*.

<sup>419</sup> F. Palazzo, M. Papa, *Lezioni di diritto penale comparato*, Giappichelli Editore, Torino, 2013, pp. 136-137.

### 3.4 In che direzione dovrebbe muoversi il programmatore per essere esente da responsabilità

Come accennato nel paragrafo introduttivo, le problematiche che nascono da situazioni riconducibili al dilemma del carrello riguardano sia aspetti di carattere prettamente giuridico, quali la possibilità di ricondurre tali ipotesi sotto la protezione offerta dallo stato di necessità, sia aspetti di carattere etico che, se messi in pratica, saranno inevitabilmente destinati a riflettere le proprie conseguenze anche sul piano giuridico.

In tale maniera, nel momento in cui le auto a guida totalmente autonoma diverranno una realtà, ipotizzando che il legislatore provveda a estendere l'ambito di applicazione dell'art. 54 c.p. o a creare una causa di giustificazione ad hoc per escludere la responsabilità del programmatore per le scelte assunte in tali situazioni, la domanda parrebbe sorgere spontanea: in quale direzione dovrebbe muoversi il programmatore per andare esente da responsabilità? Dovrebbe effettuare una programmazione informata ai valori dettati da una logica consequenzialistica o al contrario quelli portati avanti da una filosofia deontologica?

La questione appare invero di primaria importanza in quanto, osservando le modalità di funzionamento dei sistemi di machine learning, è possibile constatare come nella c.d. *decisione robotica* sono soliti entrare in gioco, non solamente i dati inseriti originariamente<sup>420</sup>, quelli forniti dai sensori e quelli acquisiti tramite l'esperienza, ma anche una serie di valori condivisi dalla collettività. Il robot dovrebbe perciò ispirare le proprie decisioni comportamentali in modo tale da salvaguardare al meglio tali valori<sup>421</sup>.

In questo senso, qualora il legislatore dovesse prendere posizione, sposando i valori promossi ad esempio dalla filosofia utilitaristica o al contrario deontologica, una programmazione degli AV rispettosa degli stessi non dovrebbe far sorgere alcuna ipotesi di responsabilità penale. Detto in altri termini; se una determinata scelta venisse normata, ossia la *metrica del valore* venisse generalmente condivisa e la macchina agisse seguendo la suddetta impostazione, ogni conseguenza derivante dalle sue decisioni, comprese

---

<sup>420</sup> A. Carcaterra, *Machinae Autonomae e decisione robotica*, cit. p. 38. Egli fa riferimento in particolare ai "modelli della realtà".

<sup>421</sup> A. Carcaterra, *Machinae Autonomae e decisione robotica*, cit. p.56.

quelle indesiderate e lesive, dovrebbero essere tollerate dall'ordinamento rappresentando per l'appunto il minimo danno sociale arrecabile<sup>422</sup>.

In simili scenari, il problema della responsabilità penale appare dunque spostarsi sulla condivisione e sulla ideazione della funzione di valore<sup>423</sup>.

Si pone così, con riferimento all'ipotesi della responsabilità per i danni derivanti da scelte pre-programmate degli AV, l'ulteriore questione circa le difficoltà nell'effettuare un bilanciamento tra beni in conflitto. Una futura ed ipotetica causa di giustificazione, il cui fondamento risiede nel mancato contrasto dell'azione con gli interessi della comunità, richiederà dunque un simile bilanciamento. Ci si chiede ora se, in sede di programmazione degli AV, sia effettivamente possibile impostare un simile bilanciamento seguendo una logica consequenzialistica, ossia una logica che consenta di salvaguardare sempre e comunque il bene o i beni dal valore sociale maggiore, ovvero se tale bilanciamento sia da considerarsi precluso a priori.

#### **A FAVORE DI UNA SOLUZIONE CONSEQUENZIALISTICA.**

Come analizzato precedentemente, è possibile immaginare, in linea di principio, scenari di emergenza che coinvolgono gli AV con caratteristiche analoghe al caso *Re A*. Si consideri ad esempio il caso in cui, in sede di programmazione, si opti per l'opzione che consenta di salvare un numero maggiore di persone modificando così la rotta seguita dall'AV e sacrificando, di conseguenza, la motociclista in transito sull'opposta carreggiata.

Si tratta di una situazione in cui il programmatore, come nel caso *Re A*, ha adottato una soluzione permeata dalla logica consequenzialistica. Così come la Corte d'Appello ha acconsentito alla separazione dei due gemelli siamesi preventivando il decesso di uno dei due, si potrebbe ipotizzare che la legge possa in futuro acconsentire a una pre-programmazione degli AV che, in scenari di emergenza, favorisca la soluzione di stampo utilitaristico.

---

<sup>422</sup> *Ivi.* p. 61.

<sup>423</sup> *Ibidem.*

In un simile scenario, si potrebbe persino insistere sul fatto che la posizione del programmatore appaia, in linea di principio, moralmente più sicura di quella dei dottori in *Re A* poiché esso ha preso la sua decisione, non rispetto ad uno scenario reale e conosciuto, ma piuttosto su scenari futuri, ipotetici e non conosciuti dettagliatamente ex ante. In sede di programmazione la scelta è stata infatti adottata facendo affidamento su alcune caratteristiche generali che si assumono sussistere in tali situazioni, non potendosi conoscere in anticipo tutti i dettagli dello scenario di emergenza: per fare un esempio, il dettaglio che la motociclista investita fosse madre lavoratrice di una famiglia molto numerosa<sup>424</sup>.

Il programmatore potrebbe pertanto appellarsi alla *dottrina del doppio effetto*<sup>425</sup> sostenendo in tal maniera che, seppur avendo chiaramente previsto che dalla sua condotta sarebbe derivata la morte di qualcuno, non si abbia ucciso intenzionalmente nessuno: l'uccisione intenzionale è infatti cosa diversa dal risolvere, in senso consequenzialistico il dilemma del carrello; in più, si consideri che "intenzionale" significa "deliberatamente diretto verso una o più persone individuali"<sup>426</sup>.

Si precisa tuttavia che, la dottrina del doppio effetto, salverebbe probabilmente il programmatore esclusivamente dall'accusa di omicidio intenzionale, non anche da quella di omicidio colposo o persino di omicidio con dolo eventuale<sup>427</sup>.

Molti filosofi e laici trovano, in definitiva, che sia del tutto irrazionale la diffusa riluttanza ad accettare l'introduzione, nella legge, di un semplice ragionamento utilitaristico applicato a omicidi commessi in situazioni di emergenza e del tutto imprevedibili. Harris, a sostegno di un'opzione utilitaristica, ha sottolineato che - seppur l'obiettivo della morale e della legge sia quello di proteggere alcuni beni di base, tra cui la vita umana - accettare l'idea per cui, nel momento in cui risulti impossibile proteggere tutte le istanze di quel bene, la vita di tutti, si dovrà essere autorizzati, se non addirittura obbligati, a cercare di salvare quante più vite possibile<sup>428</sup>.

---

<sup>424</sup> Santoni de Sio, F. *Killing by Autonomous Vehicles and the Legal Doctrine of Necessity*. cit. 416 e ss.

<sup>425</sup> Per un approfondimento sul tema si veda ad esempio W. Quinn, *Azioni, intenzioni e conseguenze: la dottrina del doppio effetto*, in «Philosophy and Public Affairs», 18, 1989, pp.334 e ss.

<sup>426</sup> G. C. Christie, *The defense of necessity considered from the legal and moral points of view*, in «Duke Law», 48, 1999, pp. 975 e ss.

<sup>427</sup> F. Santoni de Sio, F. *Killing by Autonomous Vehicles and the Legal Doctrine of Necessity*. cit. p. 417.

<sup>428</sup> J. Harris, *The survival lottery*, in «Philosophy» 50, 1975, pp. 81 e ss.

Tuttavia, una semplice lettura consequenzialistica del principio del "male minore" nei casi di omicidio rimane piuttosto impopolare. Pertanto risulta doveroso, ai fini di un'indagine approfondita, comprendere le ragioni concettuali ed etiche alla base dello scetticismo avvertito nei confronti di una simile soluzione.

#### **CONTRO UNA SOLUZIONE CONSEQUENZIALISTICA.**

Una delle ragioni più importanti che si cela dietro lo scetticismo nei riguardi di un calcolo utilitaristico in uno scenario di emergenza è rappresentato dal problema dell'*incommensurabilità*, problema che può essere sottoposto a tre diverse letture: può essere infatti letto come problema concettuale, epistemico o morale:

- sulla scorta della lettura concettuale di tale problema, si rileva in particolare la difficoltà di effettuare una misurazione ragionevole del valore dei diversi beni coinvolti e potenzialmente sacrificabili in un simile scenario di emergenza. Se già può risultare difficile stabilire un equilibrio tra il valore di beni differenti quando sono in gioco solamente beni materiali, la difficoltà è destinata ad aumentare esponenzialmente quando vengono in rilievo vite umane diverse: poiché ogni vita è diversa, e poiché la perdita di essa di solito si riflette sulla condizione di altre persone, il loro valore non risulta misurabile sulla base di meri standard oggettivi e di conseguenza risulterebbe impossibile stabilire a priori che, per esempio, cinque vite sono più preziose di una vita. In definitiva, secondo la lettura concettuale, tale valore è determinato da valutazioni soggettive, non è possibile rilevare standard o parametri oggettivi per effettuare misurazioni affidabili<sup>429</sup>.
- un secondo aspetto relativo al problema dell'*incommensurabilità* è messo in risalto da una lettura epistemica dello stesso. Anche qualora, in linea di principio, possano essere riscontrati degli standard oggettivi per decidere se una determinata vita è più preziosa di altre, si rileva che gli agenti coinvolti in circostanze dilemmatiche potrebbero non avere le informazioni, il tempo o la capacità di effettuare una valutazione adeguata<sup>430</sup>. Inoltre, come messo in evidenza dal Professor Ian Dennis, non va sottovalutata la difficoltà nel valutare il peso di quelle conseguenze a lungo termine e meno visibili dei danni causati dal fatto di consentire, anche se solo in simili scenari, il mancato rispetto

---

<sup>429</sup> F. Santoni de Sio, F. *Killing by Autonomous Vehicles and the Legal Doctrine of Necessity*, cit., p. 418

<sup>430</sup> *Ivi*.

di una norma generale che vieta il danneggiamento intenzionale di interessi di base come la vita<sup>431</sup>. Tale osservazione rivela in realtà una problematica più generale: a differenza degli esperimenti del pensiero filosofico con riferimento a occasionali giudizi morali, quale il dilemma del carrello, la legge non dovrebbe solo considerare la desiderabilità delle conseguenze del diverso corso delle azioni in una singola occasione, ma dovrebbe anche riflettere sugli effetti a lungo termine nella società di quelle regole e di quelle politiche che consentono, o addirittura incoraggiano, determinati tipi di comportamenti<sup>432</sup>.

- infine, la lettura normativa, che riflette considerazioni di carattere deontologico, fa leva sul fatto che, i diritti fondamentali quale quello alla vita non potrebbero e non dovrebbero essere confrontati o scambiati con altri diritti: devono solo essere rispettati in quanto tali.

Si noti come la maggior parte degli aspetti messi in evidenza dal problema dell'incommensurabilità rimangono tali anche se trasferiti sul piano di pre-programmazione degli AV e questo induce a rifuggire, in tale sede, dall'adozione di una logica prettamente utilitaristica.

Si potrebbe in realtà sostenere che il prodotto dell'intelligenza artificiale dovrebbe essere in grado di superare il problema dell'incommensurabilità qualora venga preso in considerazione esclusivamente dal punto di vista epistemico. In linea di principio, e ipotizzando di aumentare le attuali capacità di calcolo degli agenti artificiali<sup>433</sup>, potrebbe infatti essere inclusa nella programmazione di un veicolo una valutazione di fattori quali l'età, la professione, le condizioni sanitarie e finanziarie, lo stato familiare e le relazioni sociali delle persone coinvolte in un potenziale incidente, in modo molto simile a quanto avviene nelle valutazioni effettuate dalle compagnie assicurative sulla disponibilità e sui costi delle assicurazioni per i diversi clienti. Tuttavia, la valutazione degli effetti a lungo termine o di quelli negativi di una simile politica rimarrebbe, anche in simili scenari, un problema da non sottovalutare: si pensi a tal proposito all'impatto negativo di una simile

---

<sup>431</sup> I. Dennis, *On necessity as a defence to crime: possibilities, problems and the limits of justification and excuse*, in «Criminal Law and Philosophy», 1, 2009, pp. 29 e ss.

<sup>432</sup> F. Santoni de Sio, *Killing by Autonomous Vehicles and the Legal Doctrine of Necessity*, cit., p. 418.

<sup>433</sup> Si tratta, ancora una volta, di uno scenario non realistico dato l'attuale stato della tecnologia, ma non concettualmente impossibile nel lungo periodo.

valutazione sulla percezione della sicurezza o sulla vita delle persone sistematicamente svantaggiate dalla politica utilitaristica<sup>434</sup>.

Al contrario, una lettura concettuale o morale del problema dell'incommensurabilità pone questioni etiche e legali non superabili nemmeno in via concettuale: nessun ipotetico aumento nell'acquisizione di informazioni o nel potere di calcolo da parte dei sistemi artificiali sarebbe in grado di risolvere le problematiche sollevate. Il diritto individuale alla vita non dipende dalle caratteristiche empiriche oggettive riscontrabili nei diversi beni individuali o nelle persone<sup>435</sup>.

In definitiva, si noti come la legge esistente<sup>436</sup>, in linea con quanto sopra riferito con riguardo al problema dell'incommensurabilità, pone rigidi limiti all'esecuzione di semplici calcoli utilitaristici per giustificare il danneggiamento intenzionale di interessi di base come la proprietà, e prima ancora la vita umana, in scenari di emergenza. Ciò costituisce un dato da non sottovalutare nel momento in cui si sonda il terreno di una futura regolamentazione degli scenari di emergenza che vedranno coinvolti gli A.V.

#### **L'ESIGENZA DI RISPETTARE IL DOVERE DI NON DANNEGGIARE E PROTEGGERE**

Da quanto analizzato sopra si ricava che, al fine di impostare correttamente la questione della responsabilità nel caso di danni provocati dalle scelte dei programmatori, bisognerà rifuggire dall'adozione integrale un approccio sia di stampo prettamente utilitaristico che prettamente deontologico nella risoluzione di scenari riconducibili al dilemma del carrello. Come spesso avviene, la risposta dovrà essere ricercata nel mezzo. Così come l'approccio utilitaristico non potrà essere preso in considerazione in ogni eventualità, stante la problematica dell'incommensurabilità, ci si dovrà chiedere se anche il rigoroso rispetto del dovere delle società produttrici di non danneggiare terze parti e quello di proteggere la vita dei propri clienti sarà destinato, in specifiche situazioni, ad essere limitato.

Come spiegato all'inizio di tale capitolo, ci troviamo infatti nella *società delle mangrovie*, un'ambiente in cui diversi diritti, doveri e visioni filosofiche si scontrano, dando così luogo a quell'ambiente salmastro ed ibrido in cui solo le mangrovie sono in grado di

---

<sup>434</sup> F. Santoni de Sio, *Killing by Autonomous Vehicles and the Legal Doctrine of Necessity*, cit., pp. 419 e ss.

<sup>435</sup> *Ivi*.

<sup>436</sup> Quando si parla in generale di "legge" si fa riferimento agli aspetti comuni delle leggi dei diversi ordinamenti, non solamente alla legge italiana.

crescere. Una futura regolamentazione della materia dovrà infatti assumere e dotarsi di quelle caratteristiche che la rendano in grado di conciliare istanze diverse e spesso in contrasto tra loro.

Partendo dall'analisi del dovere di non danneggiare intenzionalmente o negligenemente terze parti si noti come la legge già prevede attualmente un elevato dovere di diligenza in capo agli automobilisti stante il fatto che essi gestiscono "armi potenzialmente pericolose". Sulla base di ciò risulta pertanto ragionevole presumere che un onere simile, o persino un onere più elevato, dovrebbe essere imposto anche in capo alle società che producono, programmano o commercializzano AV. Tra le *terze parti* risultano tuttavia comprese anche le parti che abbiano tenuto un comportamento negligente, così ad esempio il pedone che attraversa distrattamente la strada nonostante il semaforo rosso. Il rigoroso rispetto di tale dovere sembrerebbe pertanto rendere inaccettabile, sin dalla fase della progettazione, qualsiasi suggerimento volto a indirizzare, in uno scenario di emergenza, il meccanismo di machine learning a danneggiare intenzionalmente quelle terze parti colpevoli di essersi poste in una posizione pericolosa, come nell'esempio sopra citato, i pedoni che attraversano col semaforo rosso, i ciclisti fuori dalla pista ciclabile o motociclisti senza casco.

Ci si potrebbe a tal proposito chiedere se, nel tentativo di ridurre il numero e la gravità delle vittime, sarebbe coerente con tale dovere di diligenza programmare le vetture in modo da colpire preferibilmente: a) altri veicoli rispetto a pedoni o ciclisti e b) veicoli con protezioni di sicurezza più elevate rispetto a quelle meno sicure<sup>437</sup>. Una simile programmazione appare essere a prima vista in linea non solo con il dovere di non danneggiare terze parti, ma anche con considerazioni di stampo utilitaristico in quanto riducono il rischio di incidenti mortali e condurrebbero pertanto all'obiettivo di salvare il numero maggior di vite. Tuttavia, si potrebbe anche sottolineare che, se tradotto in una politica sistematica, questo potrebbe risultare nel lungo periodo anti-utilitaristico: una simile politica potrebbe infatti scoraggiare gli investimenti e l'uso di veicoli sicuri contribuendo così a rendere nel lungo periodo più pericolosa la strada e maggiore il numero di vittime di incidenti stradali.

---

<sup>437</sup> M. Maurer, J. C. Gerdes, B. Lenz, H. Winner, *Autonomous Driving, Technical, Legal and Social Aspects*, Springer, Berlino, 2015, pp 69 e ss.

Per quanto riguarda invece il dovere posto in capo alle società di proteggere la vita dei propri clienti e il suo rapporto con l'obbligo di non danneggiare intenzionalmente parti terze, ci si potrebbe chiedere se quest'ultimo sia così rigoroso da richiedere l'uccisione dei passeggeri dell'AV quando questa risulta essere l'unica opzione alternativa aperta.

Una norma che fornisse alle case automobilistiche la "licenza" di colpire intenzionalmente un pedone innocente al fine di salvare il passeggero del veicolo sarebbe in contrasto con il dovere di non danneggiare terze parti. Allo stesso tempo tuttavia le case automobilistiche non dovrebbero essere autorizzate o addirittura incentivate a programmare i loro veicoli in modo tale da sacrificare i loro passeggeri, poiché ciò sarebbe in contrasto con il concorrente obbligo delle società di proteggere la vita dei propri clienti.

Sembrerebbe allora prospettarsi, anche in una simile situazione, la conclusione sopra raggiunta con riferimento ai danni derivanti da un malfunzionamento, quella di incentivare i produttori di AV e i responsabili politici a progettare veicoli, infrastrutture e indirizzare le impostazioni legislative del traffico stradale in modo da evitare sia le uccisioni di parti terze sia quella dei passeggeri degli AV; ad esempio progettando infrastrutture e norme in modo tale da rendere altamente improbabile la realizzazione di uno scenario così tragico e dilematico<sup>438</sup>.

Pertanto, e ritornando all'analogia con la società delle mangrovie prospettata dal filosofo Floridi, così come solo le mangrovie sono in grado di crescere in quell'ambiente salmastro dato dall'incontro tra acque dolci e acque salate, sarà necessario creare quelle condizioni che attribuiscono alla regolamentazione della materia caratteristiche simili a quelle possedute dalle mangrovie, caratteristiche che la rendano in grado di sopravvivere nonostante la confluenza di diritti e doveri in contrasto tra loro. Bisognerà quindi non solo ridisegnare la legislazione penale in materia, ma anche creare un contesto tale in cui tale legislazione possa effettivamente funzionare: la legge da sola non basta ma sarà altresì necessario un intervento relativo alle infrastrutture, alla consapevolezza delle persone, alla progettazione dei veicoli ecc.

---

<sup>438</sup> F. Santoni de Sio, *Etica e auto a guida autonoma: un libro bianco sull'innovazione responsabile nei sistemi di guida automatizzati*, Ministero olandese delle infrastrutture e dell'ambiente Rijkswaterstaat, 2016.

Vi è tuttavia una situazione riconducibile al cd. *dilemma del carrello* in cui è possibile adottare una soluzione che segua una logica utilitaristica, e che allo stesso tempo rispetti sia il dovere delle società produttrici di non danneggiare terze parti che quello di proteggere la vita dei propri clienti: situazioni in cui vengono in gioco da un lato, la proprietà di un bene, dall'altro delle vite umane.

In simili situazioni le critiche sollevate a una scelta utilitaristica da una lettura concettuale dell'incommensurabilità vengono infatti meno, in ogni ordinamento il valore della vita umana supera quello della proprietà di un bene.

In tali casi infatti può essere lecito, se non addirittura obbligatorio, danneggiare intenzionalmente la proprietà privata allo scopo di salvare vite umane<sup>439</sup>.

### L'ATTUALE SOLUZIONE

Tale linea di pensiero è stata recepita anche nelle linee guida etiche per le self driving car stilate dal ministero dei Trasporti tedesco. In realtà molti dei filosofi e degli esperti impegnati nella redazione del citato report ritengono la discussione vertente intorno al dilemma del carrello un falso problema, se non addirittura una inutile distrazione. Nelle suddette linee guida si legge infatti che *“Non è possibile decidere ogni volta e soprattutto a priori se salvare una vita piuttosto che un'altra, perché si tratta di una valutazione che dipende da situazioni molto specifiche, che sono influenzate anche da comportamenti imprevedibili e che, di conseguenza, non possono essere chiaramente standardizzate o programmate”*<sup>440</sup>.

Sicuramente, in determinate situazioni, è possibile, se non doveroso effettuare una valutazione ex ante: in situazioni che impongono una scelta tra il salvataggio di una vita umana e l'integrità di una vita animale o di un oggetto, la priorità va sempre data alle vite umane<sup>441</sup>.

Le vite umane tuttavia devono essere sempre considerate come tali, senza effettuare tra le stesse una distinzione in base al valore di ciascuna *“in caso di incidenti inevitabili,*

---

<sup>439</sup> Con riferimento ai droni militari autonomi viene poi in gioco un'altra ipotesi in cui i prodotti dell'intelligenza artificiale possono essere programmati in modo tale da seguire una logica utilitaristica: situazioni in cui vengono in gioco la vita di migliaia di persone e la vita di una sola persona.

<sup>440</sup> Federal Ministry of Transport and Digital Infrastructure, *Ethic commission, Automated and Connected Driving*, Report, Giugno 2017.

<sup>441</sup> *Ivi*, p. 11 pt 7

*classificare il passeggero in base a età, salute fisica o mentale, sesso e altro sarà vietato*<sup>442</sup>. Il fatto quindi che un passeggero o un pedone abbia 90 anni o 20, che abbia famiglia o meno, che sia impiegato o disoccupato, non deve in alcun modo influire sulle decisioni del software.

Ma tali linee guida prevedono un ulteriore principio di importanza fondamentale il quale, a seconda della visione personale di ciascuno può essere considerato o come un retaggio di una realtà mai sperata o, al contrario, fortemente innovativo all'interno di un'epoca di forte cambiamento e in cui le nuove tecnologie stanno assumendo un ruolo sempre più preponderante: il principio secondo cui l'uomo deve sempre avere, in ultima analisi, il potere di sottrarre il comando alla macchina.

In definitiva, e come sopra accennato, secondo gli esperti coinvolti nella redazione di tale report la questione centrale non riguarda tanto la soluzione del dilemma del carrello in quanto tale, quanto piuttosto l'evitare a priori una situazione in cui le alternative risultano essere una peggio dell'altra. Nel caso delle self driving cars, per esempio, potrebbe essere utile decidere di non sviluppare mai, sempre ammettendo che l'attuale stadio delle tecnologie lo renda possibile, le auto autonome di livello 5: quelle in cui l'uomo è solo un ospite. Il controllo ultimo deve sempre restare nelle mani dell'uomo, che deve in qualunque momento poter spegnere tutto.

Riprendendo il discorso aperto nell'introduzione del paragrafo, sembrerebbe dunque che, al momento sia stato individuato quel valore condiviso a livello istituzionale e dunque inevitabilmente destinato a riflettersi nella c.d. *decisione robotica*<sup>443</sup>. Si tratta di un valore che non affonda le sue radici né nella filosofia utilitaristica né in quella deontologica, ma piuttosto in un pregiudizio di natura antropocentrica che, allo stato, appare tutt'ora insuperabile. Per questo motivo ad oggi, l'unica scelta che consentirebbe al programmatore di andare esente da responsabilità penale in situazioni riconducibili al *trolley problem* sarebbe, paradossalmente, quella di non implementare mai nell'algoritmo di machine learning alcun tipo di scelta qualora si presentino simili scenari, ma piuttosto quella di lasciare al caso simili decisioni o, più drasticamente, quella di non giungere mai a sviluppare vetture totalmente autonome.

---

<sup>442</sup> *Ivi*.

<sup>443</sup> Si fa riferimento all'insegnamento di A. Carcaterra, *Machinae autonoma e decisione robotica*, cit., riportato a p. 156. di questo lavoro.

Si potrebbe sempre obiettare che una simile soluzione condurrebbe, nel lungo periodo, a conseguenze non del tutto positive, stante il fatto che, come evidenziato da numerosi studi, l'avvento delle auto a guida autonoma condurrebbe a una drastica riduzione del numero di incidenti stradali, ma si tratta di un'obiezione fisiologica in quanto, *quando si parla di scelte etiche, non si vince mai 1 a 0, semmai si vince 2 a 1.*

Il principio guida è pertanto uno solo: la decisione presa dall'algoritmo dev'essere sempre eticamente giustificabile; se poi questa dovesse provocare conseguenze tragiche, significa solo che, anche l'intelligenza artificiale allo stato più evoluta, non poteva farci nulla e il controllo, in ultima analisi dovrà essere relegato all'uomo.

Una simile soluzione sarebbe quindi necessaria per recuperare, almeno in parte, quella centralità della figura dell'uomo che è stata sempre di più messa da parte nel corso delle quattro rivoluzioni descritte da Floridi<sup>444</sup>.

---

<sup>444</sup>L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, cit., pp. 99 e ss.

## CONCLUSIONI

Dall'analisi svolta si può trarre la conclusione che è fortemente auspicabile, se non necessario, che al più presto venga individuata una soluzione normativa volta a regolare la questione della responsabilità penale per i danni provocati dai prodotti dell'A.I., ma che questa allo stato non può essere definitiva.

Infatti, ancor prima che la realtà tecnologica descritta divenga di uso comune, sarà doveroso, nell'ottica di un mantenimento della competitività economica all'interno del panorama internazionale, che il legislatore assuma un chiaro indirizzo politico atto a regolare la questione. Tuttavia, così come è emerso nel corso della trattazione, e trattandosi inoltre di una materia in costante evoluzione, sarebbe altresì opportuno prendere atto di come ogni soluzione proposta sia a livello istituzionale che dottrinale sarà inevitabilmente soggetta a modifiche con il mutare della realtà tecnologica e, ancor prima, sociale. Si è insistito infatti più volte sul concetto secondo il quale, dinnanzi a realtà fortemente innovative e potenzialmente pericolose quale quella in esame, si è soliti riscontrare una tendenziale ritrosia della società all'accettazione delle stesse; di conseguenza, partendo dall'assunto secondo cui le leggi sono il riflesso della società e dei suoi costumi, la normazione della materia sarà inevitabilmente influenzata dalla percezione che di essa, e del pericolo da essa rappresentato, ne avrà il tessuto sociale.

Preso atto di ciò non dovrà sorprendere come, per ogni diversa ipotesi di responsabilità esaminata, non si giungerà mai ad un'unica e solida conclusione; si indicherà piuttosto il punto d'arrivo auspicabile e le diverse soluzioni normative intermedie che con grande probabilità il legislatore, di pari passo con il sentire sociale, sarà portato a percorrere.

Nel trarre le conclusioni, si è deciso di procedere ripercorrendo i due grandi filoni in cui si sono distinte le ipotesi di responsabilità, partendo così dal caso in cui i danni abbiano tratto origine da un malfunzionamento, qui inteso come un'evoluzione anomala dell'algoritmo, e distinguendo, al suo interno, le ipotesi di responsabilità a seconda del livello di automazione del veicolo.

Per quanto riguarda le vetture semi autonome si è ritenuto di operare un'ulteriore distinzione tra quelle qualificate con un livello inferiore al 4 e quelle che invece raggiungono tale grado di automazione.

Per le prime si può definitivamente concludere che, in tali casi, il soggetto responsabile possa essere individuato nella figura del conducente umano, ancora gravato di compiti di sorveglianza attiva e, all'occorrenza, dell'onere di un intervento manuale in situazioni di pericolo.

Maggiori difficoltà insorgono invece sul terreno della regolamentazione della responsabilità penale nel caso in cui, a provocare un danno, sia una vettura dotata del quarto livello di automazione. In tali casi si potrebbe ancora prospettare che, a rispondere penalmente del danno provocato, sia il conducente c.d. *potenziale*, relegato in tali situazioni a compiti di mera sorveglianza passiva.

Si giungerebbe così alla responsabilizzazione del conducente secondo un modello imputativo omissivo per omessa sorveglianza, ossia per essere venuto meno all'obbligo di sorvegliare costantemente che il veicolo autonomo non incorra in errori o malfunzionamenti. Soluzione questa che, sebbene non priva di criticità, sia sul piano della colpevolezza che per ragioni ordine pratico, rimane tutt'ora la più accreditata in dottrina e, allo stato, quella più idonea a conciliare vantaggi sociali e istanze precauzionistiche.

Apparirebbe infatti irrealistico supporre l'introduzione, se non altro nel breve periodo, di una qualche limitazione legislativa della punibilità dei conducenti delle auto semi-autonome anche se prossime alla completa automazione, stante l'accertato scetticismo sociale ad abbandonare la figura del conducente quale *parafulmine capace di attirare su di sé le colpe per qualunque evento, indipendentemente dalla sussistenza di un reale, ragionevole ed esigibile potere di controllo degli eventi*<sup>445</sup>: riflesso della tendenziale ritrosia della società ad accettare il cambiamento.

È risultato sicuramente più complicato percorrere il terreno della regolamentazione della responsabilità nel caso in cui il malfunzionamento da cui ha tratto origine il danno si sia

---

<sup>445</sup> A.Cappellini, *Profili penalistici delle self-driving cars*, cit., p. 337 con riferimento a S.Beck, *Google Cars, Software Agents, Autonomous Weapons Systems –New Challenges for Criminal Law*, cit., pp. 171-193.

generato all'interno di un veicolo completamente autonomo, veicolo nel quale non è più prevista la presenza di una figura chiamata a svolgere compiti di sorveglianza, nemmeno in forma meramente passiva.

In tali ipotesi, in accordo con la dottrina maggioritaria, si ritiene di dover escludere la tesi sostenuta dal penalista israeliano Gabriel Hallevy, il quale afferma l'opportunità di responsabilizzare e sanzionare direttamente il robot per i danni da esso provocati. Sebbene fortemente innovativa e meticolosamente argomentata, essa appare essere una strada difficilmente percorribile non tanto per i numerosi pregiudizi di natura antropocentrica di cui la dottrina penalistica risulta essere intessuta, ma soprattutto in quanto si è alla presenza di argomentazioni dirompenti con i principi e gli assiomi cardine del diritto penale.

Si consideri inoltre che, nonostante l'asserita familiarità della proposta di Hallevy con la disciplina relativa alla responsabilità penale degli enti, che per la prima volta ha abbattuto l'assioma secondo cui l'uomo sia l'unico destinatario del diritto penale, risulta invece sussistere tra esse una differenza di fondo tale da impedirne una pacifica assimilazione e tale dunque da far ricadere la responsabilità penale direttamente in capo al robot.

Maggiormente accreditata appare dunque la proposta alternativa relativa a un'estensione della categoria penalistica della responsabilità da prodotto difettoso anche a quei danni provocati da un'evoluzione anomala del software tramite cui gli agenti artificiali operano.

L'estensione appare sicuramente plausibile con riferimento a quelle situazioni in cui il danno sia stato generato da un errore rilevabile ed individuabile con certezza; ad esempio nel caso in cui in fase di addestramento del sistema sia stato inserito un set di dati incompleto o pregiudizievole in grado di incidere sul comportamento futuro tenuto dalla macchina ovvero, nel caso in cui i vertici aziendali, dopo essere venuti a conoscenza di un problema di malfunzionamento del software, non abbiano immediatamente provveduto a porvi rimedio. In tali casi si ritiene che la responsabilità debba ricadere sul soggetto incaricato a gestire l'area di rischio in cui il difetto ha avuto origine, ossia quel soggetto che non abbia previsto un evento che avrebbe ragionevolmente dovuto prevedere, ovvero che abbia mal implementato nel sistema le modalità con cui il mezzo stesso dovrebbe reagire al presentarsi di specifici dati concreti.

Il nodo realmente problematico nella risoluzione della questione si è tuttavia presentato con riferimento a quei casi in cui l'evento lesivo tragga origine da un errore non riconoscibile. In tali ipotesi l'estensione della categoria penalistica del danno da prodotto difettoso non è apparsa pacifica in quanto, quei già discutibili canoni individuati dalla giurisprudenza per l'imputazione del danno da prodotto, risultano qui inadatti e insufficienti a reggere l'addebito dell'evento.

Si è constatato come, una volta scomparsa la figura del conducente ad assorbire le colpe dei sinistri stradali, la tensione politica tra vantaggi sociali e istanze precauzionistiche sia destinata a presentarsi nella sua nudità. E' per questo motivo che si è dovuto affrontare nel corso della trattazione un dilemma, anche etico, di difficile risoluzione: autorizzare e di conseguenza accettare in taluni casi un vuoto di tutela ovvero al contrario proibire radicalmente tale nuova tecnologia, perdendo così i benefici da questa apportati anche in termini di vite umane.

Nel trattare la questione non si è tuttavia potuto prescindere dall'affrontare il tema relativo alla percezione del rischio tecnologico da parte del tessuto sociale, il quale, anche se in maniera meno visibile, è in grado di esercitare una forte influenza sullo sviluppo della regolamentazione della materia.

Analizzando quindi congiuntamente istanze sociali e problematiche di carattere penalistico si è giunti alla conclusione che, almeno in un primo periodo, residuerà per ancora un certo lasso temporale una forte resistenza autorizzare veicoli completamente driverless, in linea con le severe regole stabilite in tema di produzione e distribuzione di prodotti potenzialmente pericolosi, che la dottrina maggioritaria ritiene applicabili anche ai prodotti dell'A.I.

Si è tuttavia rilevato che, qualora la circolazione delle stesse venga effettivamente autorizzata, sia probabile l'adozione di un indirizzo legislativo e giurisprudenziale orientato verso l'imputazione, nei confronti di chi immette nel mercato agenti intelligenti, di ogni tipo di danno possibile, anche se prevedibile esclusivamente in modo solo astratto e generico, ovvero anche in assenza di leggi scientifiche consolidate e corroborate da studi o infine, addirittura nella totale assenza di ogni spiegazione causale.

Si ritiene dunque incardinarsi un'ipotesi di responsabilità per colpa in capo al programmatore o al produttore per la mancata previsione di qualsivoglia evento dannoso che costituisca il logico sviluppo delle modalità di funzionamento dell'agente artificiale da lui programmato, anche qualora non siano noti tutti gli anelli del processo causale. Resterebbe in tal maniera una componente di rischio ineliminabile che ricadrebbe in capo alla società produttrice.

Si è tuttavia notato, come tale soluzione sembrerebbe non tener conto di uno dei principi cardine dell'intero diritto penale: il principio della personalità della responsabilità penale. Il programmatore si troverebbe in definitiva nella condizione di dover rispondere per fatto altrui, il fatto proprio del robot.

Per questo motivo appare auspicabile che, in un'ottica di evoluzione e sviluppo a livello non solo economico ma anche sociale, nel lungo periodo l'indirizzo legislativo, in linea con l'opinione pubblica, si orienti verso una soluzione che, se allo stato potrebbe apparire fortemente dirompente, con lo sviluppo della tecnologia da un lato, e il rafforzarsi della sua presenza nella vita quotidiana dall'altro, potrebbe rappresentare la strada ideale da seguire: ritenere che *l'uomo dietro la macchina* non sia mai responsabile per negligenza, non essendo del tutto prevedibile l'agire di un robot che abbia la capacità di elaborare informazioni in modo indipendente ed autonomo, e quindi, di assumere conseguentemente decisioni altrettanto autonome.

Seguendo tale impostazione il rischio tecnologico, divenendo una fonte di rischio generalmente tollerata, sarà così destinato a ricadere esclusivamente sulla vittima alla stregua di quanto avviene in caso di calamità naturali, nonostante la natura degli stessi sia ben diversa. Il rischio tecnologico infatti, seppur non prevedibile, è stato generato dall'uomo e sempre dall'uomo è stato tollerato: l'eventuale futura circolazione di vetture dotate del quinto livello di automazione sarà frutto di un giudizio di valore che stimi socialmente prevalente l'interesse a giovare dei benefici derivanti dall'uso della tecnologia driverless rispetto agli eventuali ed esigui danni che l'A.I potrebbe causare.

È altrettanto auspicabile, tuttavia, che il legislatore, al fine di apportare una tutela adeguata seppur non onnicomprensiva, si muova su diversi versanti: primo tra tutti quello relativo alle infrastrutture e, in secondo luogo ma non meno importante, operando sul piano risarcitorio. Una volta razionalizzata a livello antropologico e culturale l'esistenza

del suddetto vuoto di responsabilità, l'esigenza di soddisfare le legittime pretese delle vittime potrebbe essere infatti colmata su piani differenti da quello penalistico, tra cui per l'appunto quello risarcitorio, informato a schemi imputativi che, in ipotesi eccezionali, consentono di prescindere dai principi di responsabilità per fatto proprio e colpevolezza, principi che al contrario rappresentano la colonna portante della responsabilità penale.

Considerazioni diverse sono state invece svolte con riferimento al secondo dei due filoni in cui si è deciso di strutturare l'analisi delle diverse ipotesi di responsabilità, ossia i casi in cui l'evento lesivo sia stato causato da una scelta effettuata in sede di programmazione, in un momento quindi antecedente rispetto quello in cui il danno si è verificato. Si sono prese in considerazione esclusivamente quelle scelte assunte in situazioni riconducibili al c.d. *trolley problem*, situazioni eticamente dilemmatiche nelle quali in capo al programmatore viene necessariamente imposto il sacrificio di una o più vite al fine di salvarne altre.

La difficoltà insita nella risoluzione della questione risiede nella circostanza secondo la quale la qualifica di prodotto autonomo modifica sostanzialmente i termini entro cui si dovrà inquadrare il problema della responsabilità in simili situazioni. Nelle automobili tradizionali, controllate esclusivamente dall'azione umana, tale scelta verrà infatti adottata direttamente dall'automobilista in una situazione di pericolo estremo e soprattutto attuale, comportando così, in sede di giudizio penale, l'eventuale applicazione della scriminante prevista dall'art. 54 c.p. relativa allo *stato di necessità*.

Al contrario, in un'automobile a guida autonoma, la scelta che la stessa implementerà sarà una decisione standardizzata e pre-programmata, circostanza questa che induce ad escludere, stante l'assenza dell'attualità del pericolo, la possibilità di applicare la detta causa di esclusione della responsabilità, comportando di conseguenza numerose problematiche sul piano giuridico, oltre che sicuramente sul piano etico.

In simili scenari, si è ritenuto di escludere l'applicabilità dell'art. 54 c.p., almeno così come attualmente regolato, sia nel caso in cui esso venga qualificato alla stregua di una scusante sia nel caso in cui venga inquadrato tra le cause di giustificazione. Nel primo caso infatti, intendendosi per scusante una causa di esclusione della colpevolezza nei confronti di chi abbia agito sotto l'effetto di una pressione psicologica tale da rendere inesi-

gibile un diverso comportamento rispettoso dei precetti penali, appare evidente l'inconciabilità con la situazione in cui versa il programmatore, il quale viene a trovarsi a decidere, lontano dalla situazione di pericolo, sulla vita e la morte di soggetti non conosciuti e non identificabili a priori.

Allo stesso modo, si è ritenuto di dover escludere l'applicabilità dell'art. 54 c.p. anche sposando l'opposta dottrina che qualifica lo stato di necessità alla stregua di una causa di giustificazione, ossia situazioni nelle quali un fatto di regola vietato dall'ordinamento, a seguito di un bilanciamento tra diversi beni in conflitto, viene consentito o addirittura imposto dall'ordinamento, risultando di conseguenza esente da pena in quanto non più considerato antiggiuridico. Seguendo tale impostazione, sebbene detto bilanciamento, in via ipotetica, potrebbe essere richiesto al programmatore chiamato a decidere su situazioni riconducibili al c.d. *dilemma del carrello*, risulterebbe comunque assente l'elemento dell'attualità del pericolo, circostanza che, ancora una volta, induce a escludere la possibilità di applicare lo stato di necessità quale causa di esclusione della responsabilità del programmatore.

Si ritiene inoltre che l'inapplicabilità dell'art. 54 c.p. alla situazione in esame debba comunque essere ribadita nonostante la legge e la giurisprudenza, con particolare riferimento a quella anglo-americana, abbiano in casi determinati già consentito alcuni danni intenzionali, addirittura alcuni omicidi, dopo misurati esercizi di giudizio e quindi in assenza dell'attualità del pericolo. Si tratta infatti di decisioni relative a casi la cui particolarità e la cui unicità ne impediscono l'estensione a regola generale.

Si ipotizza dunque che una delle strade potenzialmente percorribili in sede istituzionale sia quella relativa alla previsione di una causa di giustificazione ad hoc diretta a deresponsabilizzare il programmatore per le scelte adottate in simili scenari. Percorrendo questa strada sarà dunque il legislatore che, tramite canali istituzionali, provveda a stabilire la direzione in cui il programmatore dovrebbe muoversi nel caso in cui si trovi a dover regolare situazioni in cui la causazione di un danno si presenti come inevitabile, ossia provveda a stabilire se il programmatore debba implementare una scelta fondata su considerazioni di carattere utilitaristico o al contrario deontologico.

Tuttavia, e condividendo le linee guida presentate dal ministero dei trasporti tedesco, la maggior parte degli Stati uniformati ai valori e ai principi etici europei ritiene di non poter

in alcun modo rilasciare alcuna *licenza di uccidere*, e legittimare così la possibilità di decidere a priori se salvare delle vite piuttosto che altre, trattandosi di valutazioni dipendenti da situazioni molto specifiche, del tutto inidonee ad essere standardizzate o pre-programmate. Si ritiene in definitiva di dover lasciare la questione ancora aperta in quanto fortemente soggetta ai futuri sviluppi della tecnologia che, ci si augura, riuscirà a colmare un vuoto che a livello etico si presenta invalicabile.

## BIBLIOGRAFIA

- G. Abbà, *Felicità, vita buona e virtù. Saggio di filosofia morale*, Las, Roma, 1995.
- K. Alhelt, *The Applicability of the EU Product Liability Directive to Soft-ware, Comparative and International Law*, in «Journal South Africa», 2, 2001, pp. 188-209.
- G. E. Anscombe, *Modern Moral Philosophy*, in «Philosophy», 124, 1958, pp. 1-19.
- F. Antolisei, *Manuale di Diritto Penale, Parte Generale*, XVI ed., Giuffrè, Milano, 2003.
- K. A. Appiah, *Experiments in Ethics*, Harvard University Press, Cambridge, 2010.
- P. Asaro, *Un corpo da calciare, ma ancora nessuna anima da dannare: prospettive legali sulla robotica*, in Robot Ethics a cura di P. Lin, K. Abney, G. Bekey, MIT Press, Cambridge-Massachusetts, 2012, pp. 169-186.
- I. Asimov, *Io Robot*, Mondadori, Milano, 2018.
- J. L. Austin, *Una richiesta di scuse*, in *Philosophical Papers*, a cura di J. O. Urmson e G. J. Warnock, Clarendon Press, Oxford, 1961.
- C. Bagnoli, *Teoria della responsabilità*, Il Mulino, Bologna, 2019.
- F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in «Diritto penale e Uomo», 5, 2019, pp. 1-33.
- S. Beck, *Google Cars, Software Agents, Autonomous Weapons Systems –New Challenges for Criminal Law*, in *Robotics, Autonomics and the Law*, a cura di E. Hilgendorf, U. Seidel, Nomos, Baden-Baden, 2017, pp. 227-251
- U. Beck, *La società del rischio, verso una seconda modernità*, Carocci Editore, Roma, 2000.
- R. E. Bellman, *Applied Dynamic Programming*, Princeton-Oxford, 2010.
- J. Bentham, *Deontologia*, a cura di S. Cremaschi, La Nuova Italia, Firenze 2000.
- G. Bettiol e L. Pettoello Mantovani, *Diritto Penale, Parte Generale*, XII ed., CEDAM, Padova, 1986.
- G. Bettiol, *Sul diritto penale dell'atteggiamento interiore*, in «Rivista italiana di diritto e procedura penale», 1971.
- V. Bettiol, *Scritti giuridici, Vol II*, Cedam, Padova, 1966.

- H. Blumenberg, *Naufragio con spettatore. Paradigma di una metafora dell'esistenza*, Il mulino, Bologna, 2001.
- J. F. Bonnefon., A. Shariff, I. Rahwan, *The social dilemma of autonomous vehicles*, in «Science», 352, 2016, pp. 1573-1576.
- R. Borsari, *Intelligenza artificiale e responsabilità penale: prime considerazioni*, in «Rivista di diritto dei media», 3, 2019, pp. 262-268.
- A. Cappellini, *Profili penalistici delle self-driving cars*, in «Diritto Penale Contemporaneo», 2, 2019, pp. 325-353.
- M. Cappuccilli, *Platone, Tutte le opere*, L'Erma di Bretschneider, Roma, 2003.
- A. Carcaterra, *Machinae autonomae e decisione robotica*, in *Decisione Robotica*, a cura di A. Carleo, Il Mulino, Bologna, 2019, pp. 33-59.
- D. Castronuovo, *La colpa penale*, Giuffrè, Milano, 2009.
- D. Castronuovo, *Responsabilità da prodotto e struttura del fatto colposo*, in «Rivista Italiana di Diritto e Procedura Penale», 1, 2005, pp. 301-340.
- M. Ceresa-Castaldo, *Procedura Penale delle Società*, Giappichelli Editore, Torino, 2017.
- G. C. Christie, *The defense of necessity considered from the legal and moral points of view*, in «Duke Law», 48, 1999, pp. 975-1042.
- F. Collet, *Deep learning con Python. Imparare a implementare algoritmi di apprendimento profondo*, Apogeo, Milano, 2020.
- G. Contissa, F. Lagioia, e G. Sartor, *The Ethical Knob: ethically-customisable automated vehicles and the law*, in «Artificial Intelligence and Law», 25, 2017, pp. 365-378.
- R. Crisp, M. Slote, *Virtue Ethics*, Oxford University Press, Oxford, 1997.
- V. Cuffaro, *Codice del consumo*, Giuffrè Editore, Milano, 2012.
- E. D'Anna, G. Valle, A. Mazzoni, I. Strauss, et al., *A closed-loop hand prosthesis with simultaneous intraneural tactile and position feedback*, in «Science Robotics», 27, 2019, pp. 1-11.
- C. Darwin, *L'origine della specie*, Rizzoli, Segrate, 2009.
- A. De Mauro, *Big data analytics: guida per iniziare a classificare e interpretare dati con il machine learning*, Apogeo, Milano, 2019.

- T. De Mauro, *Grande dizionario italiano dell'uso*, Torino, 2000.
- I. Dennis, *On necessity as a defence to crime: possibilities, problems and the limits of justification and excuse*, in «Criminal Law and Philosophy», 1, 2009, pp. 29-49.
- P. Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, New York, 2013.
- F. Douma, e S. Palodichuk, *Criminal Liability Issues Created by Autonomous Vehicles*, Santa Clara Law Review, 4, 2012, pp. 1157-1169.
- M. Duni, *Limiti all'obbligo di prevedere le imprudenze altrui*, in «Rivista Giuridica della Circolazione e dei Trasporti», 4, 1964, pp. 317-343.
- R. Endsley, D. J. Garald, *Theoretical underpinnings of situation awareness: A critical review*, in «Situation awareness analysis and measurement», 1, 2000, pp. 3-21.
- G. Fiandaca, *Diritto penale. Parte generale*, VII ed., Zannichelli, Roma, 2019.
- G. P. Fletcher, *Rethinking criminal law*, Oxford University Press, New York, 2000.
- L. Floridi, J.W. Sanders, *On the Morality of Artificial Agents*, in «Minds and Machines», 3, 2004, pp. 349-379.
- L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Raffaello Cortina Editore, Milano, 2017.
- L. Floridi, *Soft Ethics e la governance del digitale*, in «Philosophy and Technology», 1, 2018, pp. 1-8.
- L. Floridi, *Digitalisation, opportunities and challenges*, Chalmers University of Technology, Gothenburg, Sweden, March 2017.
- P. Foot, *The Problem of Abortion and the Doctrine of the Double Effect*, 1967, in «Oxford Review», 5, pp. 5-15,
- L. Fridman, B. Reimer, L. Angell, S. Seaman et al., *MIT Advanced Vehicle Technology Study: Large-Scale Naturalistic Driving Study of Driver Behavior and Interaction with Automation*, in «IEEE Access», 11, 2019, pp. 102021- 102038.
- H. Fry, *Hello World. Essere umani al tempo del digitale*, Bollani Boringhieri, Torino, 2019.
- M. C. Gaeta, *Automazione e responsabilità civile automobilistica*, «in Responsabilità civile e previdenza», 5, 2016, pp. 1718-1750.

- F. Giunta, *Illiceità e colpevolezza nella responsabilità colposa*, CEDAM, Padova, 1993.
- S. Gless, E. Silverman, T. Weigend, *If robots cause harm, who is to blame? Self-driving cars and criminal liability*, in «New Criminal Law Review», 19, 2016, pp. 412-436.
- J. P. Goldschmidt, *Der Notstand, ein Schuldproblem*, Manzsche k.u.k. Hof-Verlags- und Universitäts-Buchhandlung, 1913.
- J. Habermas, *L'inclusione dell'altro, studi di teoria politica*, Feltrinelli editore, Milano, 2013.
- G. Hallevy, *When Robots Kill. Artificial Intelligence under Criminal Law*, Northeastern University Press, Boston, 2013
- G. Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, Dordrecht, 2015
- G. Hallevy, *Unmanned Vehicles: Subordination to Criminal Law under the Modern Concept of Criminal Liability*, in «Journal of Law, Information & Science», 2, 2011/2012, pp. 200-211.
- G. Hallevy, *The Criminal Liability of Artificial Intelligence Entities, from Science Fiction to Legal Social Control*, in «Akron Intellectual Property Journal», 4, 2010, pp. 171-201.
- G. Hardin, *The Tragedy of the Commons*, in «Science», 3859, 1968, pp. 1243-1248.
- J. Harris, *The survival lottery*, in «Philosophy» 50, 1975, pp. 81-87.
- H. Hart, *Il concetto di legge*. Clarendon Press, Oxford, 1961.
- Hegel, *Lineamenti di filosofia del diritto*, trad. it. a cura di F. Messineo, Laterza, Bari, 1913.
- A. Hevelke e J. Nida-Rümelin, *Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis*, in «Science and Engineering Ethics», 21, 2015, pp. 619-630.
- E. Hilgendorf, *Automated Driving and the Law*, in, *Robotics, Autonomics and the Law*, a cura di U. Seidel, Nomos, Baden-Baden, pp. 171-193.
- F. H. Hsu, *IBM's Deep Blue Chess grandmaster chips*, in «IEEE Micro », 2, 1999, pp. 70-81.
- M. Iaselli, *Risoluzione UE sulla responsabilità dei robot: sarà mai possibile equiparare il robot ad un essere umano?*, in «ANDIP, associazione nazionale per la difesa della Privacy», 14 febbraio 2018.

- G. F. Italiano, *Le sfide interdisciplinari dell'intelligenza artificiale*, in «Analisi Giuridica dell'Economia, Studi e discussioni sul diritto dell'impresa», 1, 2019, pp. 9-20.
- P. Janot, *Lex Humanoïde, des robots et des juges*, Editions Thot, Fontaine, 2017.
- H. Jonas, *The Imperative of Responsibility. In search of an Ethics for the Technological Age*, University of Chicago Press, Chicago, 1984
- D. Kahneman, *Pensieri lenti e veloci*, Mondadori, Milano, 2012.
- I. Kant, *Fondazione della metafisica dei costumi*, Editori Laterza, Bari, 2018.
- I. Kant, *La dottrina del diritto*, trad. it. con prefazione e note di G. Vidari, Studio editoriale lombardo, Milano, 1916.
- G. Kasparov, *Deep thinking: Where Machine Intelligence Ends and Human Creativity Begins*, Hodder & Stoughton, London, 2017.
- P. Koopman, *A case study of Toyota unintended acceleration and softer safety*, Carnegie Mellon University, Pittsburg, 2014.
- R. Lanctot, H. Cohen, *Accelerating the Future: The Economic Impact of the Emerging Passenger Economy*, in «Strategy Analytics », 1, 2017, pp. 1-30.
- G. Lattanzi, *Reati e Responsabilità degli Enti, guida al dlgs. 8 giugno 2001, n.231*, Giuffrè Editore, Milano, 2010.
- P. Ligutti, *I diritti come fini. La riflessione di Amartya Sen tra consequenzialismo e teorie deontologiche*, in «Filosofia politica», 3, 2004, pp. 461-480.
- W. F. Lloyd, *Two Lectures On the Checks to Population*, Nabu Press, Firenze, 2014.
- J. Locke, *Due trattati sul governo*, a cura di L. Pareyson, UTET, Torino, 2010.
- Lucrezio, *De rerum natura*, II, 1-4. Trad. it. a cura di E. Cetrangolo, Lucrezio, Della natura, Sansoni, Firenze, 1969.
- M. Turing, *Computing machinery and intelligence*, in «Mind», 59, 1950, pp. 433-460.
- M. B. Magro, *A.I.: la responsabilità penale per la progettazione, la costruzione e l'uso dei robot*, in «Pluris, quotidiano giuridico», 12 giugno 2018.
- G. Marinucci, E. Dolcini, G. Gatta, *Manuale di diritto penale. Parte generale*, VIII ed., Giuffrè, Milano, 2018.

- M. Maurer, J. C. Gerdes, B. Lenz, H. Winner, *Autonomous Driving, Technical, Legal and Social Aspects*, Springer, Berlino, 2015.
- E. Mezger, *Diritto Penale (Strafrecht)*, trad. it. dell'Avv. Mandalari, CEDAM, Padova, 1935.
- M. Molinari, *Assedio all'Occidente*, La Nave di Teseo Editore, Milano, 2019.
- Montesquieu, *Lo spirito delle leggi*, III ed., a cura di S. Cotta, UTET, Milano, 1996.
- M. S. Moore, *Atto e criminalità: la filosofia di azione e le sue implicazioni per il diritto penale*, Oxford University Press, New York, 2010.
- F. Palazzo, M. Papa, *Lezioni di diritto penale comparato*, Giappichelli Editore, Torino, 2013.
- B. Pascal, *Pensieri*, a cura di P. Serini, Einaudi, Torino, 1967.
- C. Piergallini, *Attività produttive e imputazione per colpa: prove tecniche di diritto penale del rischio*, in «Rivista Italiana di Diritto e Procedura Penale», 1997, pp. 1473-1495.
- C. Piergallini, *Danno da prodotto e responsabilità penale, profili dogmatici e politico-criminali*, Giuffrè editore, Milano, 2004.
- C. Piergallini, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale?*, in «Rivista Italiana di Diritto e Procedura Penale», 1996, pp. 354-370.
- C. Piergallini, *La responsabilità del produttore: una nuova frontiera del diritto penale?*, in «Diritto Penale e Processo», 9, 2007, pp. 1125-1141.
- G.D. Pisapia, *Istituzioni di diritto penale*, CEDAM, Padova, 1965.
- Platone, *Alcibiade primo*, trad. it. a cura di D. Puliga, Rizzoli, Segrate, 1995.
- Platone, *Dialoghi Politici e lettere*, a cura di F. Adorno, UTET, Torino, 1988.
- D. Provolo, S. Riondato, F. Yenisey, *Genetics, Robotics, Law, Punishment*, Padova University Press, Padova, 2014.
- D. Pulitanò, *Diritto penale*, Giappichelli, Torino, 2017.
- W. Quinn, *Azioni, intenzioni e conseguenze: la dottrina del doppio effetto*, in «Philosophy and Public Affairs», 18, 1989, pp. 334-351.
- E. Rich, K. Knight, *Artificial Intelligence*, McGraw-Hill, New-York, 1991

- M. Romano, G. Grasso, *Commentario Sistematico del Codice Penale*, IV ed., vol. II, Giuffrè, Milano, 2012.
- H. Rosling, *Factfulness*, Rizzoli, Segrate, 2019.
- P. E. Ross, *Is Coronavirus Speeding the Adoption of Driverless Technology?*, in «IEEE Spectrum», 17.03.2020.
- J. C. Rossi, *La guerra che verrà: le armi autonome*, in «IRIAD Review», 11, 2016, pp. 1-20.
- P. Rossi Monti, *Paragone degli ingegneri moderni e post moderni*, Il Mulino, Bologna, 1989.
- U. Ruffolo, *Intelligenza artificiale e responsabilità*, Giuffrè Editore, Milano 2017.
- E. Rulli, *Giustizia predittiva, intelligenza artificiale e modelli probabilistici. Chi ha paura degli algoritmi?*, in «Analisi Giuridica dell'Economia», 2, 2018, pp. 533-546.
- G. Samek Lodovici, *Utilità del bene: Jeremy Bentham, l'utilitarismo e il consequenzialismo*, Vita e pensiero, Milano, 2004.
- M. Sandel, *Giustizia, il nostro bene comune*, Feltrinelli editore, Milano, Giugno 2019.
- F. Santoni de Sio, *Etica e auto a guida autonoma: un libro bianco sull'innovazione responsabile nei sistemi di guida automatizzati*, Ministero olandese delle infrastrutture e dell'ambiente Rijkswater-staat, 2016.
- F. Santoni de Sio, *Killing by Autonomous Vehicles and the Legal Doctrine of Necessity*, in «Ethic Theory Moral Practice», 2, 2017, pp. 411-429.
- A. Santosuosso, C. Boscarato, F. Corleo, *Robot e diritto: una prima ricognizione*, in «Nuova giurisprudenza civile commentata», 2, 2012, pp. 494-516.
- G. Sartor, *Gli agenti software: nuovi soggetti del ciberdiritto*, in «Contratto e impresa» 2, 2002, pp. 55-87.
- L. Scarano, *La non esigibilità nel diritto penale*, Casa Editorial Libreria Humus, Napoli, 1948.
- L. Scarano, *Libera volontà e libero Arbitrio nel Diritto Penale*, Giuffrè, Milano, 1937.
- N. Schömig, V. Hargutt, A. Neukum, I. Petermann-Stock, I. Othersen, *The interaction between highly automated driving and the development of drowsiness*, in «Procedia Manufacturing», 3, 2015, pp. 6652-6659

- K. Schwab, *La quarta rivoluzione industriale*, Franco Angeli, Milano, 2016.
- L. B. Solum, *Legal Personhood for Artificial Intelligences*, in «Illinois Public Law and Legal Theory Research Papers», 4, 2008, pp. 1231-1287.
- M. Spasari, *Diritto criminale e scienza del diritto criminale, Oggetto metodo della dogmatica criminale e della criminologia*, in «Giurisprudenza italiana», 4, 1990, pp. 506-528.
- B. C. Stahl, *Information, Ethics, and Computers: The Problem of Autonomous Moral Agents*, in «Minds and Machines», 14, 2004, pp. 67-83.
- B. C. Stahl, *Responsible Computers? A Case for Ascribing Quasi-Responsibility to Computers Independent of Personhood or Agency*, in «Ethics and Information Technology», 8, 2006, pp. 205-213.
- L. Stortoni, *Angoscia tecnologica ed esorcismo penale*, in «Rivista Italiana di Diritto e Procedura Penale», 1, 2004, pp. 71-89.
- J. J. Su, D. Vasconcellos Varagas e K. Sakurai, *One Pixel attack for fooling deep neural networks*, in «IEEE Transactions on Evolutionary Computation», 5, 2018, pp. 828-841.
- T. Hobbes, *Leviatano*, trad. it. a cura di G. Micheli, RCS libri, Milano, 2011.
- G. Tamburini, *Autonomia delle macchine e filosofia dell'intelligenza artificiale*, in «Rivista di filosofia», 2, 2017, pp. 263-275.
- M. Taylor, *Self-Driving Mercedes-Benzes Will Prioritize Occupant Safety over Pedestrians*, in «Car and Driver», 7 ottobre 2016.
- M. Tegmark, *Essere umani nell'era dell'intelligenza artificiale*, Raffaello cortina editore, Milano, 2018.
- Thomson J.J. , *The Trolley Problem*, in «The Yale Law Journal», 5, 1985, pp. 1395-1415.
- Tucidide, *La guerra guerra del Peloponneso*, trad. it. a cura di F. Ferrari, Milano, Rizzoli, 1985.
- S. Turkel, *La conversazione necessaria. La forza del dialogo nell'era digitale*, Einaudi, Torino, 2016.
- P. Veneziani, *Regole cautelari "proprie" ed "improprie" nella prospettiva delle fattispecie colpose causalmente orientate*, CEDAM, Padova, 2003.
- M. Von Buri, *Über Kausalität und deren Verantwortung*, Gebhardt, Lipsia, 1873.

B. Wilson, J.Hoffman, J.Morgenstern, *Predictive Inequity in Object Detection*, in «arxiv.org», 21 febbraio 2019, pp. 1-11.

D. Wuyts, *The Product Liability Directive – More than two decades of defective products in Europe*, in «Journal European Tort Law», 5, 2014, pp. 1-34

## **NORMATIVA E GIURISPRUDENZA**

Camera dei Deputati, servizio studi-Dipartimento dei trasporti, *La mobilità e il futuro: l'auto a guida autonoma*, Dossier n° 275, 31 gennaio 2017.

C. Cost., sentenza n. 364, 24 marzo 1988.

Cass. Pen., sez. III, sentenza n. 13225, 26 giugno 2015.

Cass. pen, *Considerazioni sul principio del ne bis in idem nella recente giurisprudenza Europea: la sentenza del 4 Marzo 2014, Grande Stevens e altri contro Italia*, Rel. n. 35/2014, Roma 8 maggio 2014.

Commissione Europea, *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, L'intelligenza artificiale per l'Europa*, 237 final, Bruxelles, 25.4.2018.

Commissione Europea, *Libro Bianco Sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia*, 65 final, Bruxelles, 19.2.2020.

Commissione Europea, *Relazione della Commissione al Parlamento Europeo, al Consiglio e al Comitato Economico e Sociale Europeo sull'applicazione della direttiva del Consiglio relative al ravvicinamento delle disposizioni legislative, regolamentari ed amministrative degli Stati Membri in materia di responsabilità per danno da prodotti difettosi*, 496 final, Bruxelles, 14.9.2006

Corte eur. dir. uomo, Plenaria, caso n. 5100/71, *Engel and Others v. the Netherlands*, 8 giugno 1976

Direttiva 85/374/CEE del del 25 luglio 1985 rDirettiva 85/374/CEE del 25 luglio 1985, *Direttiva del Consiglio relativa al ravvicinamento delle disposizioni legislative, regolamentari ed amministrative degli Stati Membri in materia di responsabilità per danno da prodotti difettosi*

Dlgs., 8 Giugno 2001, n° 231, *Disciplina della responsabilità amministrativa delle persone giuridiche, delle società e delle associazioni anche prive di personalità giuridica.*

Dlgs 2 Febbraio 2001, n° 25., *Attuazione della direttiva 1999/34/CE, che modifica la direttiva 85/374/CEE, in materia di responsabilità per danno da prodotti difettosi.*

Dlgs 6 Settembre 2005, n° 206, *Codice del consumo.*

Dpr. 24 maggio 1988, n° 224, *Attuazione della direttiva CEE n. 85/374 relativa al ravvicinamento delle disposizioni legislative, regolamentari e amministrative degli Stati membri in materia di responsabilità per danno da prodotti difettosi, ai sensi dell'art. 15 della legge 16 aprile 1987, n. 183.*

Federal Minister of Transport and Digital Infrastructure, *Automated and connected driving*, report Giugno 2017.

GDPR (General Data Protection Regulation) del Parlamento Europeo e del Consiglio del 27 aprile 2016, approvato con Regolamento UE 2016/679.

Grande Stevens e altri contro Italia, Rel. n. 35/2014. Roma, 8 maggio 2014.

L. 12 aprile 2019 n° 31, *Disposizioni in materia di azione di classe.*

L. 3 maggio 2019 n° 37, *Disposizioni per l'adempimento degli obblighi derivanti dall'appartenenza dell'Italia all'Unione europea.*

Ministero dello sviluppo economico, *Piano Nazionale Industria 4.0*, 2016.

National Transportation Safety Board Office of Public Affairs, *Preliminary Report Released for Crash Involving Pedestrian*, 24 maggio 2018.

Open Letter to the European Commission Artificial Intelligence and Robotics, 05/04/2018 <http://g8fip1kplyr33r3krz5b97d1-wpengine.netdna-ssl.com/wp-content/uploads/2018/04/RoboticsOpenLetter.pdf>

Parlamento Europeo, *Auto a guida autonoma in UE: dalla fantascienza alla realtà*, in [europarl.europa.eu](http://europarl.europa.eu), 14 gennaio 2019.

Parlamento Europeo, *Raccomandazioni alla commissione concernenti norme di diritto civile sulla robotica*, Risoluzione del 16 febbraio 2017.

Parlamento Europeo, *Una politica industriale europea globale in materia di robotica e intelligenza artificiale*, Risoluzione del 12 febbraio 2019.

Queen vs. Dudley and Stephens, 14 Queens Bench Division 273, 9 dicembre 1884.

Re A (Children)(Conjoined Twins) [2001] 2 WLR 480, Court of Appeal.

SAE International, *Level of driving automation are defined in new Sae international standards J3016*, in sae.org, 2014.

Statuto di Roma della Corte Penale Internazionale, 17 luglio 1998.

The High-Level Expert Group in A.I., *Ethic Guidelines for Trustworthy A.I.*, in Reports and studies of European Commission, Aprile 2019.

World health organization, *Global status report on road safety*, 2018.

## SITOGRAFIA

AON, *Global Analytics*, 2016. <https://www.aon.com/2016techreport/default.jsp>

ASCO *Nuove correlazioni tra l'uso della tecnologia cognitiva di Watson e la cura del cancro*, in «IBM news Room», 1 giugno 2017. <https://it.newsroom.ibm.com/announcements?lnk=hmhm&l=25&o=125>

J. Brill, *Data Brokers: A Call for Transparency and Accountability*, in Reports of Federal Trade Commission, maggio 2014. <https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>

G.Coppola, *Fiat Teams With Silicon Valley's Aurora on Self-Driving Tech*, in «Bloomberg Technology», 10 giugno 2019. <https://www.bloomberg.com/news/articles/2019-06-10/flat-teams-with-silicon-valley-s-aurora-on-self-driving-vehicles>

Corporale vehicle observatory, *Le auto a guida autonoma: siamo già nel futuro?*, report ottobre 2016. [https://cms-static.arval.com/sites/default/files/it--cvo/media/pdf/libri/le\\_auto\\_a\\_guida\\_autonoma\\_pdf\\_protetto.pdf](https://cms-static.arval.com/sites/default/files/it--cvo/media/pdf/libri/le_auto_a_guida_autonoma_pdf_protetto.pdf)

L. Floridi, *Incontro con Luciano Floridi*, inserito nel progetto "Dalla maschera al robot" con il contributo di Fondazione Cariplo, 9 gennaio 2020. <https://www.youtube.com/watch?v=0E1kNRpNW10>

L. Floridi, *Onlife*, La vita al tempo del digitale, Politecnico di Milano, 5 ottobre 2019. <https://www.youtube.com/watch?v=vVp0SHqDEwY>

C. Giffi, J. Vitale, *Global Automotive Consumer Study*, Deloitte, 2018. <https://www2.deloitte.com/it/it/pages/consumer-industrial-products/articles/automotive-consumer-study---deloitte-italy---consumer---industri1.html>

D. Godau, K. Grambow, N. Gill, R. Mehl, M. Winkler, *Beyond the Car*, CapGemini, 2017. <https://www.capgemini.com/consulting-de/wp-content/uploads/sites/32/2017/05/cars-online-study-2017.pdf>

K. Hao, *Self-driving cars may be more likely to hit you if you have dark skin*, in «MIT technology review», marzo 2019. <https://www.technologyreview.com/2019/03/01/136808/self-driving-cars-are-coming-but-accidents-may-not-be-evenly-distributed/>

S. Hawking, E. Musk, S. Wozniak, N. Chomsky, et al., *An open letter to the united nations convention on certain conventional weapons*, 24th International Joint Conference on Artificial Intelligence, Buenos Aires, 2015. <https://www.stopkillerrobots.org/2015/07/aicall/>

D. La Cava, *Toyota investe 500 mln\$ in Uber, sinergie nelle auto a guida autonoma*, in «Finanza Online», Agosto 2018. <https://www.finanzaonline.com/notizie/toyota-investe-500-mln-in-uber-sinergie-nelle-auto-a-guida-autonoma>

La redazione di Forbes, *La tua prima auto a guida autonoma potrebbe essere di Amazon*, in «Forbes», 10 Febbraio 2019. <https://forbes.it/2019/02/10/amazon-investe-guida-autonoma-aurora/>

B. Marr, *The 4th Industrial Revolution Is Here - Are You Ready?*, in «Forbes», Agosto 2018. <https://www.forbes.com/sites/bernardmarr/2018/08/13/the-4th-industrial-revolution-is-here-are-you-ready/#361263f4628b>

*moralmachine.mit.edu*

J. Muller, *No Hands, No Feet: My Unnerving Ride In Google's Driverless Car*, in «Forbes», 21 marzo 2013. <https://www.forbes.com/sites/joannmuller/2013/03/21/no-hands-no-feet-my-unnerving-ride-in-googles-driverless-car/#3bd2dea1554e>

S. Pini, *Guida autonoma, Hyundai e Kia entrano in Aurora*, in «Il Sole 24ore», 15 giugno 2019. <https://www2.ilsole24ore.com/art/notizie/2019-06-15/guida-autonoma-hyundai-e-kia-entrano-aurora-095535.shtml?uuid=AC2PTWR>

I. Rahwan, *The Social Dilemma Of Driverless Cars*, in TEDxCambridge, settembre 2016. <https://www.tedxcambridge.com/talk/the-social-dilemma-of-driverless-cars/B. Simonetta, Coronavirus, l'intelligenza artificiale di Alibaba lo identifica in 20 secondi con una Tac>, in «Sole 24 ore», 6 marzo 2020. <https://www.ilsole24ore.com/art/coronavirus-l-intelligenza-artificiale-alibaba-identifica-20-secondi-una-tac-ADacdTB>

B. Simonetta, *Così big data e intelligenza artificiale stanno battendo il coronavirus in Cina*, in «Sole 24 ore», 9 marzo 2020. <https://www.ilsole24ore.com/art/coronavirus-l-intelligenza-artificiale-alibaba-identifica-20-secondi-una-tac-ADacdTB>

F. Suman, *Dilemmi morali per le auto a guida autonoma*, in «Il Bo Live Università di padova», novembre 2018. <https://ilbolive.unipd.it/it/news/dilemmi-morali-auto-guida-autonoma>

T. Worstall, *When Should Your Driverless Car From Google Be Allowed To Kill You?*, in «Forbes», 18 giugno 2014. <https://www.forbes.com/sites/timworstall/2014/06/18/when-should-your-driverless-car-from-google-be-allowed-to-kill-you/#3666c1cafa5b>