

Ugo Ruffolo
(a cura di)

Tech e-Law

Intelligenza artificiale

Il diritto, i diritti, l'etica

Prefazione di

Guido Alpa
Augusto Barbera



GIUFFRÈ FRANCIS LEFEBVRE

Sezione non inclusa

GLI AUTORI

Guido Alpa

Emerito di Diritto Civile nell'Università di Roma "La Sapienza", già Presidente del Consiglio Nazionale Forense

Andrea Amidei

Dottore di ricerca nell'Università di Bologna

Barbara Banorri

Dottore di ricerca nell'Università di Bologna

Augusto Barbera

Emerito di Diritto Costituzionale nell'Università di Bologna, Giudice della Corte Costituzionale

Roberto Bichi

Presidente del Tribunale di Milano

Marilù Capparelli

Dottore di ricerca nell'Università di Bologna e Direttore del Dipartimento Affari Legali Google

Beniamino Caravita di Toritto

Ordinario di Istituzioni di Diritto Pubblico nell'Università di Roma "La Sapienza", Vice-Presidente dell'Associazione Italiana dei Costituzionalisti

Giuseppe Corasaniti

Sostituto Procuratore Generale presso la Corte di Cassazione, docente di Informatica Giuridica nell'Università di Roma "La Sapienza"

Maria Costanza

Ordinario di Istituzioni di Diritto Privato nell'Università di Pavia

Gli autori

Lorenzo d'Avack

Emerito nell'Università Roma Tre, Presidente del Comitato Nazionale per la Bioetica

Francesco Di Giovanni

Ordinario di Diritto Privato nell'Università di Bari

Luca Enriques

Professor of Corporate Law nella Università di Oxford, già Ordinario alla Università LUISS Guido Carli e Commissario alla CONSOB

Giusella Finocchiaro

Ordinario di Diritto di Internet e di Diritto Privato nell'Università di Bologna

Alberto Maria Gambino

Prorettore dell'Università Europea di Roma e Ordinario di Diritto Privato

Vincenzo Guggino

Segretario Generale dell'Istituto di Autodisciplina Pubblicitaria

Francesca Lagioia

Senior Research Fellow nello European University Institute

Mariateresa Maggiolino

Direttore del Corso di Giurisprudenza all'Università Bocconi di Milano e Associato di Diritto Commerciale

Sandro Mainardi

Ordinario di Diritto del Lavoro nell'Università di Bologna

Rodolfo Mancini

Chief Legal Officer di SACE S.p.A.

Vittorio Manes

Ordinario di Diritto Penale nell'Università di Bologna, componente del *Group of Experts on Artificial Intelligence and Criminal Law* istituito dal *European Committee on Crime Problems* - CDPC del Consiglio d'Europa

Mariachiara Manzi

Dottore di ricerca nell'Università Europea di Roma

Paolo Moro

Ordinario di Filosofia del Diritto e Informatica Giuridica nell'Università di Padova

Ugo Pagallo

Ordinario di Filosofia del Diritto nell'Università di Torino, membro dell'*Expert Group on Liability for New Technologies - New Technologies Formation* nominato dalla Commissione Europea

Filippo Patroni Griffi

Presidente del Consiglio di Stato

Ugo Ruffolo

Già Ordinario di Diritto Civile nell'Università di Bologna

Giovanni Sartor

Ordinario di Informatica Giuridica nell'Università di Bologna

Mario Savini Nicci

Consulente giuridico generale della Fondazione Policlinico Universitario Agostino Gemelli I.R.C.C.S.

Paola Severino

Professore di Diritto Penale, Vice-Presidente della Università LUISS Guido Carli

Luca Tomassini

Presidente e Ceo del gruppo Vetrya

Alberto Torini

Coordinatore *Unit Corporate Finance Legal Affairs* di SACE S.p.A.

Antonio Uricchio

Ordinario di Diritto Tributario, e già Rettore, dell'Università di Bari "Aldo Moro", Presidente dell'ANVUR

Gli autori

Alberto Vacchi

Presidente e Amministratore Delegato di IMA S.p.A.

Giuseppe Vetrugno

Associato di Medicina Legale all'Università Cattolica del Sacro Cuore

Termine estratto capitolo

Parte I ETICA

1.**La rivoluzione tecnologica
e la nuova era digitale: problemi etici**

di Lorenzo d'Avack

Sommario: 1. Una premessa: il potere della scienza. — 2. Rivoluzione tecnologica e digitale: potenzialità e rischi. — 3. L'uso dei *big data* e degli algoritmi. — 4. I robot. — 5. Le raccomandazioni etiche. — 6. Una conclusione.

1. Una premessa: il potere della scienza

La storia dei nostri giorni e del nostro secolo può essere descritta come una 'rivoluzione tecnologica'. Una rivoluzione che suscita un senso di disorientamento in considerazione della velocità con cui le tecnologie stanno modificando il mondo conosciuto. D'altronde si assiste ad un rapido sviluppo delle tecnoscienze e delle tecnologie convergenti: combinazione e sinergia di diversi settori scientifici (nanotecnologie, ingegneria genetica, biotecnologie, tecnologie dell'informazione, scienze cognitive, neuroscienze, intelligenza artificiale, robotica, ecc.). Così a fronte di un cieco e forse voluto iniziale ottimismo nel non porre limiti alla scienza, oggi a seguito dell'enorme progresso, continuo, irreversibile della trasformazione tecnologica e tecnocratica del mondo, gli Stati nazionali, come le organizzazioni internazionali, sono alla ricerca di una regolamentazione di questa nuova rivoluzione industriale, suscettibile di toccare tutti gli strati sociali, rendendo imprescindibile che la legislazione, o meglio una *governance*, auspicabilmente globale, ne disciplini le implicazioni e le conseguenze giuridiche ed etiche.

Una *governance* la cui difficile genesi si esplica attraverso la presenza di un conflitto di valori e di diritti, esplicito o mascherato, tra le esigenze e gli interessi della ricerca e le esigenze e gli interessi dell'uomo, inteso quest'ultimo al contempo sia come individuo singolo, sia come testimone e rappresentante dell'umanità, anche per quanto concerne le generazioni future¹.

La scienza è un potere: la potenza suprema a disposizione. I grandi progetti biotecnologici, che riguardano i geni e il genoma, e la clonazione, la procreata artificiale, il potenziamento della vita, la costruzione di una vita artificiale, il rapporto uomo-macchina, sono l'opera di un potere che realizza l'utopia bio-

¹ D'AVACK, *Il dominio delle biotecnologie. L'opportunità e i limiti dell'intervento del diritto*, Torino, 2018, cap. I.

tecnologica' e che vede referenti principali gli scienziati, le industrie, lo Stato e i media. In essa si trovano le nuove fonti di profitto del terzo millennio². In una tale visione, l'uomo si eleva allo stadio di vero signore e proprietario della natura in grado di creare una 'nuova stirpe umana', giustificata dalla 'moralità del benessere'. Il bioeticista inglese John Harris preannuncia che la rivoluzione nella biologia molecolare consentirà di costruire nuove forme di vita su ordinazione e per quanto concerne il genere umano una 'nuova stirpe' (*wonderwoman* e *superman*), prodotta da interventi sulla linea germinale con l'introduzione di nuovi geni, oggi tecnicamente possibile con l'*editing-genetico*. « La pretesa degli individui modificati di costituire una stirpe (che peraltro essi potrebbero benissimo non avanzare mai) poggerebbe chiaramente, ma in modo decisivo, sul fatto che essi possono trasmettere questa nuova costituzione mediante normale riproduzione con altri membri della stessa 'stirpe' ... Naturalmente quando parlo di 'genitori', penso ai 'genitori genetici', e questi naturalmente, per essere tali, hanno bisogno soltanto di essere donatori di ovuli, di spermatozoi o magari di embrioni »³. Siamo ad una bioetica post-umana, che ai limiti di una fantascienza, ipotizza la manipolazione del corpo umano finalizzata alla creazione di ibridi uomo-macchina (il *cyborg*, mente artificiale in un corpo biologico). La considerazione della soggettività va oltre l'umano, oltre l'attuale vita organica e include nuovi soggetti artificiali, sintetici e cibernetici⁴.

Sono dunque sanciti la fine della post-modernità e l'inizio di una era finalizzata a far vivere le biotecnologie. Ne consegue per i *maîtres à penser*, favorevoli ad uno sviluppo senza limiti della scienza, anche la necessità di bandire quegli atteggiamenti considerati antichi, arcaici, portati a porre limiti morali alle nuove tecnologie. L'intero dibattito sulla legittimità o meno delle biotecnologie riguarda l'emancipazione dalla schiavitù della natura e coloro che pongono limiti difendono a spada tratta questo schiavismo. « Molti filosofi come Jürgen Habermas — osserva Peter Sloterdijk, — che erigono barriere morali alla ricerca tecnologica sono in realtà dei reazionari, pronti a conservare un presunto ed immutabile destino biologico dell'individuo. E condannano, dalla loro ultima spiaggia teologica, ogni intervento umano nell'umano⁵.

² IRTI, *Le domande del giurista e le risposte del filosofo (un dialogo su diritto e tecnica)*, in Contr. impr., 2000, 674.

³ HARRIS, *Wonderwomen e Superman. Manipolazione genetica e futuro dell'uomo*, Milano, 1997, 270.

⁴ LECOURT, *Human Post Human*, Parigi, 2003. Non mancano trasposizioni cinematografiche di queste generazioni artificiali, basti ricordare *Blade runner*, *Intelligenza artificiale* e le diverse versioni di *Matrix*.

⁵ Intervista in L'Espresso, **Termine estratto capitolo** conferenza tenuta da Sloterdijk nel 2003.

2.

Etica e diritto dell'Intelligenza Artificiale nella governance del digitale: il *Middle-out Approach*

di Ugo Pagallo

Sommario: 1. Introduzione. — 2. Sull'etica dell'IA e altre tecnologie digitali. — 3. Il diritto dell'IA e altre tecnologie digitali. — 4. Conclusioni.

1. Introduzione

Nel corso degli ultimi anni abbiamo assistito alla pubblicazione di un considerevole numero di dichiarazioni e linee guida attorno ai principi etici dell'intelligenza artificiale (IA) e di altre tecnologie digitali. Mi limito a ricordare i rapporti dell'Organizzazione per la cooperazione e sviluppo economico (OCSE), i vertici presso le Nazioni Unite su “*AI for Good*”, i cd. Principi Asilomar del *Future of Life Institute*, i documenti di OpenAI, della *Partnership on AI*, del Centro per la quarta rivoluzione industriale del *World Economic Forum*, e di altri ancora, cui si aggiungono le strategie sull'IA sul piano nazionale e sovranazionale. Per quanto riguarda l'Unione Europea, ad esempio, sono da segnalare la Risoluzione del Parlamento europeo del febbraio 2017; l'Opinione sull'IA del Comitato economico e sociale nel maggio di quello stesso anno; oltre alla Strategia sull'IA della Commissione europea dell'anno successivo e ai lavori del Consiglio d'Europa. A conferma del trend, è degno di nota che il 16 luglio 2019, quando il Parlamento europeo ha eletto la prima donna presidente della Commissione europea, Ursula von der Leyen, quest'ultima ha rilasciato le proprie linee guida politiche, in cui, al terzo paragrafo su “un'Europa all'altezza dell'era digitale”, l'argomento ruota proprio attorno alle sfide e opportunità dell'IA. Con le sue parole, l'obiettivo è nei “miei primi 100 giorni di presentare una proposta legislativa per un approccio europeo coordinato alle implicazioni etiche e umane dell'IA”.

In aggiunta alle dichiarazioni politiche e alle raccomandazioni sui principi etici dell'IA, occorre pertanto prendere in considerazione i contributi sul fronte giuridico. Significativamente, il 25 aprile 2018, la Commissione europea ha istituito ben tre Gruppi di esperti (v. doc. IP/18/3362): un primo gruppo è stato creato per mettere a fuoco i principi etici dell'IA; il secondo per studiare le eventuali modifiche alla vecchia direttiva comunitaria (1985) in materia di responsabilità da prodotto difettoso; e, infine, il terzo gruppo sul tema della responsabilità giudica extra-contrattuale per le “tecnologie emergenti” (di cui ho

avuto l'onore e il piacere di fare parte). Mentre il gruppo di esperti sul fronte etico ha avuto modo di pubblicare la prima versione dei propri lavori, ossia *Ethics Guidelines for Trustworthy AI*, nell'aprile 2019, il gruppo giuridico sulla responsabilità extra-contrattuale ha presentato il proprio rapporto, ossia *Liability for AI and other Digital Technologies*, nel novembre 2019.

Su queste basi, il presente saggio intende dunque investigare la relazione che si dà tra etica e diritto, partendo proprio dai documenti, le dichiarazioni, le linee guida e i rapporti fin qui adottati dalle istituzioni o dalle organizzazioni non governative circa le sfide normative dell'IA. Il saggio è stato diviso in due parti. Nella prima, l'indagine ruota attorno ai modi in cui il termine 'etica' può essere colto, così da precisare l'accezione più pertinente nel caso delle iniziative istituzionali e non governative menzionate poco sopra. Quando s'istituisce un gruppo di esperti sull'etica dell'IA, il riferimento va ai problemi della meta-etica o dell'etica applicata, oppure alle teorie morali? Si predilige un'etica soffice che miri a rafforzare l'odierno quadro di garanzie giuridiche, oppure quest'ultimo suggerisce un'etica *hard*, cioè a dire, una serie di principi morali per cui siamo chiamati a disattendere o soppiantare le attuali leggi giuridiche?

Nella seconda parte del saggio, si tratta invece di chiarire il tipo di diritto di cui si discute nelle iniziative istituzionali e, in genere, nel dibattito odierno sulle leggi dell'IA. Quale il rapporto tra *soft law* e *hard law*? Oltre alle norme primarie del diritto, secondo la terminologia di Herbert Hart, quale ruolo svolgono in questo caso le norme secondarie dell'ordinamento¹? Quale il rapporto tra il diritto e altri sistemi regolativi, come le forze del mercato, la tecnologia o le forze sociali, che possono entrare in competizione o anche in conflitto tra loro?

Nell'affrontare questo insieme di problemi, l'autore di queste pagine è consapevole dell'ammonimento di Benedetto Croce, il quale, nella *Riduzione della filosofia del diritto alla filosofia dell'economia* (1907), riassume gli sforzi dei filosofi del diritto di distinguere il proprio campo dalla morale con l'immagine del "Capo Horn" della scienza giuridica.

L'intento di questo scritto è più modesto. Non si mira a elucidare quale sia il rapporto che si dà tra etica e diritto, magari prendendo posizione tra le varianti del positivismo ora inclusivo ora esclusivo. Piuttosto, lo scopo è di capire come i sistemi regolativi del diritto e dell'etica s'intreccino nelle attuali discussioni degli esperti sui principi dell'IA e in rapporto a recenti iniziative istituzionali, quali quelle appena menzionate della Commissione europea, a proposito del governo del digitale.

¹ HART, *Il concetto di diritto*, trad. it. Torino, 1991.

2. Sull'etica dell'IA e altre tecnologie digitali

Parlando di etica, conviene innanzitutto distinguere tra i campi della meta-etica, delle teorie morali e dell'etica applicata². Nel caso della meta-etica, ci si occupa dei pilastri del sistema normativo, ossia, dei concetti chiave di giustizia e ingiustizia³. Nel caso delle teorie morali, il riferimento va ai modi diversi in cui i filosofi hanno definito le nozioni di obbligatorio, proibito, permesso, ecc. Così, un teorico dell'utilitarismo, o nipote di Bentham, giudicherebbe le azioni o comportamenti di un sistema d'IA in ragione dei loro esiti; un seguace di Kant, in rapporto alle intenzioni dietro a quei comportamenti o azioni; un teorico dell'etica della virtù, sulle orme di Aristotele, in virtù di ciò che rende la vita umana buona. L'etica applicata, invece, concerne i dilemmi morali che sorgono in uno specifico settore dell'esperienza. Per esempio, nel caso dei problemi etici che sorgono nel campo dell'IA, un gruppo di esperti ha sottolineato come cinque sfide siano uniche al settore dell'intelligenza artificiale, ossia il rischio di limitare il controllo umano, di rimuoverne le responsabilità, di sminuirne o deprezzarne i talenti, erodendo l'auto-determinazione umana e consentendo nuovi tipi di comportamenti impropri, o veri e propri illeciti⁴.

In rapporto a questa tripartizione di base, tra meta-etica, teorie morali ed etica applicata, bisogna poi distinguere tra *hard ethics* e *soft ethics*⁵, se si vuole, tra etica forte e soffice. Nel primo caso, la discussione etica su questioni di giustizia (meta-etica), su ciò che va proibito o permesso (teorie morali), in rapporto ai problemi unici delle tecnologie "emergenti" (etica applicata), solleva l'ulteriore questione se i diversi argomenti e principi morali debbano essere adottati anche andando contro la normativa giuridica esistente, o con buona pace di essa. Un esempio di *hard ethics* è dato dalle teorie anti-schiaviste ai tempi dell'apartheid nel Sudafrica e che si battono tuttora contro gli stati complici del fenomeno. Per converso, la *soft ethics* si pone il problema di come completare o irrobustire il quadro giuridico esistente. La funzione è analoga al ruolo della *soft law* in ambito

² PAGALLO, *When Morals Ain't Enough: Robots, Ethics, and the Rules of the Law*, in *Minds and Machines*, 2017, 27(4): 625-638.

³ RAWLS, *A Theory of Justice*, in Belknap Press of Harvard University Press, Cambridge, Mass., 1999; FLORIDI, *Toleration and the Design of Norms*, in *Science and Engineering Ethics*, ottobre 2014, 1-29; PAGALLO, *The Challenges of Digital Democracy, and How to Tackle Them in the Information Era*, in BERKICH, D'ALFONSO (a cura di), *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence*, Dordrecht, 2019, 385-393.

⁴ FLORIDI, COWLS, BELTRAMETTI, CHATILA, CHAZERAND, DIGNUM, LUETGE, MADELIN, PAGALLO, ROSSI, SCHAFER, VALCKE e VAYENA, *AI4People - An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, in *Minds and Machines*, 2018, 28(4): 689-707.

⁵ FLORIDI, *Soft Ethics and the Governance of the Digital*, in *Philosophy and Technology*, 2018, 31(1): 1-8.

giuridico. Tanto il codice etico per l'IA adottato dalla Repubblica federale tedesca nel 2019, quanto le opinioni, raccomandazioni e linee guida di Autorità Garanti nei vari settori dei dati personali, della sicurezza aerea, o nel settore dei trasporti, hanno il fine di rafforzare il diritto vigente, ora denunciandone le lacune, ora colmandole attraverso il ricorso all'interpretazione dei principi base dell'ordinamento (sia giuridico sia morale), ora rendendo le norme di quell'ordinamento più chiare o specifiche.

Ci sono fondamentalmente due ragioni per cui, quantomeno nella maggioranza dei sistemi politici occidentali, il dibattito e le iniziative istituzionali sull'etica dell'IA hanno per lo più riguardato il lato soffice dell'etica applicata (*soft ethics*), piuttosto che il lato duro della morale (*hard ethics*).

Innanzitutto, a supporto della tesi, occorre menzionare i pilastri del quadro normativo vigente, vale a dire sia la Convenzione europea dei diritti dell'uomo (CEDU) del 1950 sia la Carta dei diritti fondamentali dell'Unione europea del 2000. Le odierne discussioni sui principi etici dell'IA non hanno a che fare con la bontà dei principi solennemente proclamati in questi documenti. Piuttosto, l'obiettivo è di rafforzare i principi della Carta UE e della CEDU, stante i dilemmi, spesso inediti, dell'IA e di altre tecnologie "emergenti". I settori interessati sono dunque, principalmente, quelli dell'etica applicata e i campi del diritto elencati dalla carta e dalla convenzione: privacy, garanzie processuali, non discriminazione, ecc. Ciò non significa che, adottando il punto di vista della *soft ethics*, non esistano problemi inediti che richiedano nuovi principi. Per esempio, nel campo della tutela dei dati personali e delle garanzie processuali a tutela dell'individuo, si consideri la necessità d'introdurre un nuovo principio di *explicitability* per il trattamento dati dell'IA⁶. Tuttavia, questi nuovi principi etici devono intendersi come un modo per rinvigorire il quadro normativo esistente, non certo un modo per rivonderlo o indebolirlo.

In secondo luogo, sul piano delle fonti giuridiche, gli esperti dell'etica dell'IA si preoccupano prevalentemente dei problemi che sorgono sul fronte della legislazione ordinaria (o normativa secondaria dell'Unione europea), piuttosto che della normativa costituzionale o dei trattati internazionali. Questo non significa, ancora una volta, che la rivoluzione digitale non stia impattando alcuni cardini dell'ordinamento e, di conseguenza, che quest'ultimo ordinamento non debba essere emendato. Tra gli esempi più rilevanti, a conferma della tesi, si pensi alla necessità di integrare, poniamo, la Costituzione italiana, affiancando alla tutela

Termine estratto capitolo

3.

Macchine come noi.

Natura e limiti della soggettività robotica

di Paolo Moro

Sommario: 1. Macchine come noi. — 2. Macchine intelligenti. — 3. Macchine morali. — 4. Macchine emotive. — 5. Macchine inconsapevoli.

1. Macchine come noi

Lo sviluppo della robotica è giunto ad un momento cruciale nell'epoca contemporanea, nella quale il recente successo degli studi e delle applicazioni dell'intelligenza artificiale sta conducendo gli studiosi più avvertiti a riflettere sui limiti etici dei sistemi cibernetici¹.

In particolare, emergono oggi con sempre maggiore forza alcuni problemi giuridici che riguardano l'interazione umana con i sistemi robotici evoluti, i quali appaiono sempre più capaci di emulazione autonoma delle facoltà soggettive proprie del soggetto umano, del quale influenzano il modo di vivere ma dal cui significativo controllo restano sempre più indipendenti².

La realizzazione di macchine androidi altamente organizzate, dotate di automazione e capaci di emulare ragionamenti, comportamenti e perfino sentimenti tipicamente umani, non è più soltanto fervida immaginazione letteraria³, ma fruttuoso esperimento scientifico. Un robot sofisticato e dotato di intelligenza artificiale è oggi un sistema meccatronico complesso, che imita non soltanto ragionamenti, come la comprensione del linguaggio naturale, l'apprendimento automatico o la rappresentazione della conoscenza, ma anche comportamenti, come la capacità di decidere o la reazione a stimoli esterni, e in taluni casi anche emozioni, attraverso le espressioni del volto sintetico o altre azioni comunicative⁴.

¹ FLORIDI, COWLS, BELTRAMETTI, CHATILA, CHAZERAND, DIGNUM, LUETGE, MADELIN, PAGALLO, ROSSI, SCHAFER, VALCKE, VAYENA, *AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, in *Minds & Machines* 28(4), 2018, 689.

² YANG, WOOLDRIDGE (eds), *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, Buenos Aires, Argentina, 25-31 July 2015, AAAI Press, Palo Alto, California, USA, in <http://ijcai.org/papers15/contents.php>.

³ DICK, *Do Androids Dream of Electric Sheep?*, New York, 1968.

⁴ MORO, *Biorobotics and fundamental rights. Issues and limits of artificial intelligence*, in PROVOLO, RIONDATO, YENISEY (a cura di), *Genetics Robotics Law Punishment*, Padova, 2014, 517.

È giunto il tempo delle macchine come noi⁵: l'informatica e la robotica hanno permesso di costruire *robot* che possiedono non solo l'abilità di calcolare, ossia di elaborare dati in base al modello computazionale come ogni *computer*, ma anche la capacità di decidere, attraverso la retroazione (*feedback*) o l'apprendimento automatico (*machine learning*), e anche, in certe condizioni, la facoltà di manifestare percezioni simili alle emozioni.

Questa rappresentazione meccanica dell'anima umana, riflessa come in uno specchio da un sistema non soltanto materiale (*hardware*) ma anche immateriale (*software*), sempre più simile alla persona e sempre più capace di interagire con essa, ripropone non solo la questione filosofica del discusso paragone tra coscienza umana e intelligenza artificiale, ma introduce molteplici questioni di roboetica e di tecnodiritto, che riguardano la natura e i limiti dei sistemi esperti di intelligenza artificiale⁶.

Tra questi problemi, anche per l'esperienza giuridica, assume peculiare interesse la possibilità di ipotizzare che il *robot*, presuntivamente dotato di facoltà artificiali analoghe, se non superiori, a quelle umane, possa essere titolare di una specie di autodeterminazione individuale e, quindi, anche di diritti e di doveri⁷. Infatti, la capacità di apprendimento del robot evoluto è interattiva (*interactive learning*) e incrementa progressivamente la propria autonomia operando con l'essere umano⁸.

Inoltre, lo sviluppo scientifico ed operativo delle macchine intelligenti sta portando ad applicazioni industriali della tecnologia robotica che, per realizzarsi, tendono ad eliminare le barriere tra uomo e macchina, con l'esigenza di elaborare norme etiche e giuridiche che regolino gli inevitabili conflitti generati dall'interazione tra entità sempre meno distinte tra loro.

L'attualità non solo etica, ma soprattutto giuridica, della questione è inconfutabile.

La Risoluzione del Parlamento europeo del 16 febbraio 2017, recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica, contiene un esplicito invito al massimo organo legislativo europeo, e indirettamente gli Stati membri dell'Unione, « a esplorare, esaminare e valutare, nell'ambito della valutazione d'impatto del suo futuro strumento legislativo, le impli-

⁵ McEWAN, *Machines like me*, Londra, 2019; CASTELFRANCHI, STOCK, *Macchine come noi. La scommessa dell'intelligenza artificiale*, Bari, 2000.

⁶ MORO, SARRA (a cura di), *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, Milano, 2017.

⁷ MORO, *Libertà del robot? Sull'etica delle macchine intelligenti*, in BRIGHI, ZULLO (a cura di), *Filosofia del diritto e nuove tecnologie. Prospettive di ricerca tra teoria e pratica*, Atti del XXIV Congresso della Società Italiana di Filosofia del Diritto, Roma, 2015, 525.

⁸ GOODRICH, SCHULTZ, *Human-Robot Interaction: A Survey*, in *Foundations and Trends in Human-Computer Interaction*, 1, 3, 2007, 203.

cazioni di tutte le soluzioni giuridiche possibili, tra cui: (...) l'istituzione di uno status giuridico specifico per i robot nel lungo termine, di modo che almeno i robot autonomi più sofisticati possano essere considerati come persone elettroniche responsabili di risarcire qualsiasi danno da loro causato, nonché eventualmente il riconoscimento della personalità elettronica dei robot che prendono decisioni autonome o che interagiscono in modo indipendente con terzi ».

Il documento legislativo del Parlamento Europeo è di evidente interesse per individuare il principio giuridico, peraltro tratto da una citazione letteraria, che risolva le immaginate e immaginabili controversie tra uomini e robot, sebbene non costituisca un insieme di norme positive: nelle premesse giustificative, infatti, la Risoluzione considera formalmente le quattro leggi di Asimov, tra le quali la fondamentale "legge zero" ("Un robot non può recar danno all'umanità e non può permettere che, a causa di un suo mancato intervento, l'umanità riceva danno")⁹, « come rivolte ai progettisti, ai fabbricanti e agli utilizzatori di robot, compresi i robot con capacità di autonomia e di autoapprendimento integrate, dal momento che tali leggi non possono essere convertite in codice macchina ».

La possibilità di attribuire una personalità giuridica ai robot particolarmente evoluti e sempre più simili agli umani non tanto nel *soma* quanto soprattutto nella *psiche* non è più una fantasia letteraria, ma è già acquisita nel dibattito scientifico più avveduto, nel quale si applicano categorie che tradizionalmente identificano la tutela del soggetto umano, come i diritti soggettivi, ai sistemi di intelligenza artificiale¹⁰.

Questi interrogativi, che riguardano anche il dibattito scientifico sulla cosiddetta *interaction human-robot*, possono essere affrontati riflettendo criticamente non tanto sulla tecnologia del sistema cibernetico che oggi è denominato genericamente "intelligenza artificiale" e che dipende dalla logica matematica e informatica, quanto sulla concezione antropologica che la presuppone.

La possibilità di ammettere un paragone tra uomo e macchina appare tuttora ipotizzata da molti studiosi nelle scienze naturali, morali e sociali, che di regola sostengono prospettive riferibili all'odierno postumanesimo¹¹, sulla base di un approccio dogmatico, fondato su una visione riduttiva e discutibile della soggettività umana, già presente nel razionalismo moderno, ma che tuttavia sembra ancora piuttosto diffusa nella cultura scientifica dell'epoca contemporanea.

Utilizzando una tecnologia sempre più efficiente e una scienza sempre più abile, l'uomo contemporaneo sta tentando di ricostruire un'anima artificiale nelle

⁹ ASIMOV, *The evitable conflict in I, robot*, New York, 1950.

¹⁰ KOOPS, HILDEBRANDT, JACQUET-CHIFFELLE, *Bridging the Accountability Gap: Rights for New Entities in the Information Society?*, in *Minnesota Journal of Law, Science & Technology*, 2010, 497.

¹¹ PEPPERED, *The Posthuman Manifesto*, in *Kritikos*, 2, 2005, 1-16.

macchine¹², dotate talvolta di un corpo sintetico, ma sempre più spesso immateriali e definite da un software.

È un processo ormai evidente verso un obiettivo preciso, che la scienza persegue nei sistemi meccatronici attraverso una sorta di rispecchiamento delle principali facoltà della psiche umana, che già Platone aveva elaborato in una notissima e, come spesso succede con i grandi filosofi, sostanzialmente invariata tripartizione tra la ragione, la volontà e la passione¹³.

Avvertendo che l'armonia delle tre facoltà spirituali è il riflesso della giustizia che ciascuno realizza nelle relazioni sociali, Platone aveva raccomandato di tenere insieme queste parti dell'anima e non di disgiungerne la struttura unitaria per comprendere correttamente la soggettività umana: al contrario, con una singolare sequenza diacronica, l'uomo tecnologico e postmoderno si sta dedicando a potenziare nelle macchine informatiche e robotiche prima l'intelligenza, poi la volontà e infine l'emozione, in base ad una concezione individualista e frammentata dell'uomo, concepito come persona artificiale ad un'unica dimensione¹⁴.

2. Macchine intelligenti

Le macchine intelligenti nascono quando matematici e informatici del secolo scorso considerano appropriato comparare la natura dei processi mentali di elaborazione e apprendimento delle informazioni alla struttura biologica dei processi naturali e sociali dell'attività fisiopsichica.

Il progetto computazionale di costruzione di un'intelligenza meccanica risale indubbiamente ad Alan Turing, il quale inventa l'odierno *computer*, analizzando il processo razionale del calcolo combinatorio e riferendosi con significativo linguaggio metaforico agli "stati mentali" per definire le configurazioni interne della macchina e alla costruzione di un "cervello" per delineare l'esecuzione della sua idea¹⁵.

Successivamente, Norbert Wiener collega espressamente la ricerca sugli esseri viventi a quella sulle macchine per garantire un adeguato controllo su entrambi e sostiene l'impossibilità teorica di distinguere il sistema nervoso umano da un sistema di comunicazione meccanico come quello telefonico, determinando il

¹² GALLINA, *L'anima delle macchine. Tecnodestino, dipendenza tecnologica e uomo virtuale*, Bari, 2015.

¹³ PLATONE, *Repubblica*, 439 a-441 c, in *Opere*, vol. II, Bari, 1967, 262-265.

¹⁴ MARCUSE, *One-Dimensional Man: On the Society of Advanced Industrial Society*, Boston, 1964.

4.

Le decisioni algoritmiche tra etica e diritto

di Giovanni Sartor e Francesca Lagioia

Sommario: 1. Il concetto di intelligenza artificiale. — 2. L'Intelligenza Artificiale nel nuovo millennio. — 2.1. Intelligenza artificiale generale e specifica. — 2.2. Modelli logici e apprendimento automatico. — 2.3. Metodi per l'apprendimento automatico. — 3. Apprendimento automatico e dati personali. — 4. Intelligenza artificiale e *big data*: opportunità e rischi. — 5. Le decisioni algoritmiche tra equità e discriminazione. — 6. IA, valori e norme giuridiche. — 6.1. Un quadro etico per l'intelligenza artificiale. — 6.2. Principi e norme giuridiche. — 6.3. Gli interessi da proteggere. — 7. Conclusioni.

1. Il concetto di intelligenza artificiale

Nella sua definizione più ampia, *l'Intelligenza Artificiale* (IA) può essere definita come il tentativo di creare macchine « capaci di eseguire attività che richiedono intelligenza quando svolte da esseri umani »¹. L'intelligenza artificiale comprende un ampio insieme di ricerche e tecnologie diverse, accomunate dall'obiettivo di realizzare sistemi artificiali capaci di comportamenti intelligenti. Non è una singola tecnologia, ma piuttosto un vasto insieme di approcci e tecnologie, che a livelli diversi e in modi differenti mostrano un comportamento intelligente in numerosi e svariati contesti. Di recente, il gruppo di esperti ad alto livello sull'IA (AI HLEG), istituito dalla Commissione europea, ha elaborato una più ampia definizione di IA, secondo cui: « I sistemi di intelligenza artificiale (IA) sono sistemi software (e verosimilmente anche hardware) progettati da esseri umani che, dato un obiettivo complesso, agiscono all'interno di una dimensione fisica o digitale, percependo il loro ambiente attraverso l'acquisizione di dati, interpretando i dati raccolti, siano essi strutturati o non strutturati, ragionando sulla conoscenza, o elaborando le informazioni derivate da questi dati e selezionando tra tutte le azioni possibili le migliori per raggiungere l'obiettivo indicato. I sistemi di intelligenza artificiale possono utilizzare regole simboliche o apprendere un modello numerico e possono anche adattare il loro comportamento analizzando il modo in cui l'ambiente è influenzato dalle loro azioni passate »².

¹ Si vedano KURZWEIL, *The Age of Intelligent Machines*, Cambridge, 1990, 14, e RUSSELL, NORVIG, *Artificial Intelligence. A Modern Approach*, Upple Saddle River, 2016, sez. 1.1.

² AI-HLEG, High-Level Expert Group on Artificial Intelligence, *A definition of AI: Main capabilities and Scientific Disciplines*, 2019.

Questa definizione può essere accolta purché si tenga a mente che la maggior parte dei sistemi d'intelligenza artificiale esegue solo una parte delle attività in essa elencate: il riconoscimento di *pattern* (ad esempio, il riconoscimento di immagini di piante o animali, volti o espressioni umane), l'elaborazione del linguaggio naturale (ad esempio, la comprensione di lingue diverse, la traduzione da una lingua all'altra, i metodi per combattere lo spam o la capacità di rispondere a quesiti), la capacità di dare suggerimenti pratici (ad esempio, raccomandare i prodotti e i servizi da acquistare, ottenere informazioni, eseguire una pianificazione logistica o ottimizzare i processi industriali), ecc. D'altro canto, alcuni sistemi di IA possono combinare molte di queste capacità, come nell'esempio dei veicoli a guida autonoma o dei robot militari e di assistenza.

Il termine "algoritmo" è spesso usato per identificare applicazioni d'intelligenza artificiale, ad esempio attraverso locuzioni come "decisioni algoritmiche". Il concetto di algoritmo ha in realtà un carattere più generale, poiché include qualsiasi sequenza di istruzioni definite in modo univoco per eseguire efficacemente un'attività, in particolare, ma non solo, attraverso calcoli matematici³. Gli algoritmi informatici possono essere molto semplici, specificando, ad esempio, come mettere in ordine alfabetico elenchi di parole o trovare il massimo comune divisore di due o più numeri; oppure essere molto complessi, come ad esempio gli algoritmi per la crittografia o la compressione di file digitali, il riconoscimento vocale o le previsioni finanziarie. Ovviamente, non tutti gli algoritmi riguardano o si occupano d'intelligenza artificiale, ma tutti i sistemi di intelligenza artificiale — come qualsiasi sistema informatico — presuppongono l'uso di algoritmi.

Gli algoritmi che si occupano di IA possono applicare diversi metodi di ragionamento epistemico o pratico sui dati disponibili (ad esempio per il rilevamento di schemi e forme, l'applicazione di regole, la previsione e la pianificazione), nonché differenti metodi di apprendimento⁴. In quest'ultimo caso, un sistema di IA può essere in grado di automigliorare le proprie prestazioni, sviluppando nuove euristiche (strategie di successo che hanno prodotto buoni risultati in passato), modificando i propri dati interni o perfino generando nuovi algoritmi, come nel caso degli algoritmi genetici. Per esempio, un sistema d'intelligenza artificiale per il commercio elettronico può applicare sconti ai consumatori che soddisfano determinate condizioni (applicando regole), fornire raccomandazioni (ad esempio imparando e utilizzando correlazioni tra le caratteristiche degli utenti di una piattaforma e le loro propensioni all'acquisto), ottimizzare la gestione degli stock (ad esempio evolvendosi e implementando le migliori strategie di trading). Sebbene un sistema d'intelligenza artificiale sia costituito dal-

³ HAREL, FELDMAN, *Algorithmics: The Spirit of Computing*, Boston, 2004.

⁴ Secondo RUSSEL, NORVIG, cit., 693, "un agente sta imparando se esso migliora le proprie prestazioni rispetto a compiti futuri dopo aver compiuto osservazioni sul mondo".

l'insieme di numerosi algoritmi, esso può anche essere visto come un singolo algoritmo complesso, originato dalla combinazione di tutti gli algoritmi che svolgono le varie funzioni cui il sistema è preposto, nonché dall'algoritmo principale che orchestra tali funzioni, attivando i relativi algoritmi di livello inferiore. Ad esempio, un bot (*robot software*) che risponde a domande in linguaggio naturale sarà costituito da una combinazione orchestrata di algoritmi per rilevare suoni, acquisire strutture sintattiche, recuperare conoscenze pertinenti, fare inferenze, generare risposte, ecc.

In un sistema capace di apprendere la componente più importante non sarà l'insieme degli algoritmi appresi, che eseguono direttamente l'attività assegnata al sistema (ad esempio classificazione, previsione o decisione), quanto piuttosto gli algoritmi di apprendimento, che modificano i primi in modo che possano migliorare lo svolgimento delle proprie funzioni. Ad esempio, in un sistema di classificazione che riconosce le immagini attraverso una rete neurale, l'elemento cruciale è l'algoritmo di apprendimento (il *trainer*) che modifica la struttura interna della rete neurale (il classificatore appreso), modificandone le connessioni e i pesi, in modo che la rete neurale possa classificare correttamente gli oggetti di interesse (ad es. animali, suoni, volti e atteggiamenti).

2. L'Intelligenza Artificiale nel nuovo millennio

L'intelligenza artificiale ha attraversato ricorrenti momenti di euforia e depressione, fin dai suoi inizi negli anni '50: a previsioni eccessivamente ottimistiche sono seguite profonde disillusioni (i cosiddetti inverni dell'intelligenza artificiale)⁵. Tuttavia, è fuor di dubbio che l'Intelligenza Artificiale abbia ottenuto grandi successi negli ultimi anni. Da un lato, essa si è data una solida base scientifico-culturale: il nucleo originario costituito da informatica, ingegneria, matematica e logica è stato esteso con modelli e concetti provenienti da altre discipline come la statistica, l'economia, le neuroscienze, la filosofia e il diritto. Dall'altro lato, si è realizzata un'ampia gamma di applicazioni di successo, che sono entrate nella quotidianità dei singoli individui e della società, oltre che nelle attività economiche ed amministrative: la ricerca di informazioni (i motori di ricerca); il riconoscimento di voci, immagini e volti; la diagnostica medica; la traduzione automatica; l'analisi e la classificazione di documenti; la risposta a quesiti in linguaggio naturale (*question-answering*); i giochi, dai videogiochi agli scacchi e al go; le transazioni commerciali ad alta velocità (*high-speed trading*); i robot industriali e domestici; i veicoli autonomi in cielo, terra, e mare; l'ottimizzazione dei processi industriali e commerciali; gli assistenti virtuali; ecc.

⁵ NILSSON, *The Quest for Artificial Intelligence*, Cambridge, 2010.

2.1. Intelligenza artificiale generale e specifica

Nel campo dell'intelligenza artificiale, la ricerca persegue tradizionalmente due diversi obiettivi: l'intelligenza artificiale generale, anche detta intelligenza artificiale forte, e l'intelligenza artificiale specifica, detta anche intelligenza artificiale debole. L'intelligenza artificiale generale si propone di realizzare sistemi capaci di esibire la maggior parte delle capacità cognitive umane, allo stesso livello degli esseri umani o anche a un livello superiore (la cosiddetta superintelligenza)⁶. L'intelligenza artificiale specifica si propone obiettivi più modesti, cioè la costruzione di sistemi capaci di eseguire singoli compiti, o aspetti di essi, che normalmente richiedono intelligenza.

La futura creazione di un'intelligenza artificiale generale già oggi suscita serie preoccupazioni: un'intelligenza artificiale generale potrebbe migliorare sé stessa con progressione esponenziale, trasformandosi in breve tempo da umana in superumana, e acquisendo, grazie alla sua intelligenza superiore, capacità tali da diventare incontrollabile⁷. L'umanità verrebbe a trovarsi in una posizione di inferiorità simile a quella degli animali rispetto all'uomo, cosicché la vita degli individui e la stessa esistenza della nostra specie finirebbe per dipendere dalla volontà e quindi dalla benevolenza delle entità superintelligenti. Alcuni scienziati e tecnologi (come Steven Hawking e Bill Gates) hanno affermato la necessità di anticipare il « rischio esistenziale »⁸ rappresentato dalla superintelligenza artificiale adottando misure che prevengano l'emergere di entità superintelligenti o almeno assicurino che tali entità siano « amichevoli » verso la specie umana, rispettandone i valori. Altri studiosi, invece, vedono con favore il momento in cui l'umanità darà origine ad un'intelligenza capace di superare l'umanità stessa. Un'intelligenza artificiale capace di automigliorare sé stessa rappresenterebbe la « singolarità » in grado di accelerare lo sviluppo della scienza e della tecnologia, con risultati che potrebbero non solo risolvere i problemi dell'umanità (povertà, sottosviluppo, ecc.), ma anche consentire il superamento dei limiti biologici dell'uomo (malattia, invecchiamento, ecc.), e diffondere l'intelligenza nel cosmo⁹.

I rischi legati all'emergere di una "intelligenza generale artificiale" non devono essere sottovalutati: si tratta infatti di un problema molto serio che porrà grandi sfide in futuro. Per quanto le opinioni degli scienziati a riguardo siano divergenti, la maggior parte di essi ritiene che tale obiettivo sarà raggiunto entro la fine

⁶ BOSTROM, *Superintelligence*, Oxford, 2014.

⁷ BOSTROM, cit. Tale possibilità era già stata anticipata da TURING, *Intelligent Machinery, A Heretical Theory*, in *Philosophia Mathematica* 4 ([1951] 1996).

⁸ PARKIN, *Science fiction may be the best model of AI*, in *AI Obsessions*, in *The Guardian*, 14 June 2015.

Parte II
DIRITTO CIVILE: RESPONSABILITÀ,
CONTRATTI, PERSONA E *PRIVACY*

Sezione I A.I. E RESPONSABILITÀ

Termine estratto capitolo

2. Intelligenza Artificiale e responsabilità da prodotto

di Andrea Amidei

Sommario: 1. L'autonomia dell'A.I. ed i correlati problemi di riparto delle responsabilità. — 2. Le prospettive di revisione della Direttiva *Product Liability*. — 3. L'A.I. come “prodotto” e l'algoritmo quale suo componente. — 4. L'integrazione tra “prodotto” e “servizio” e le responsabilità del *trainer* e del fornitore di *training data*. — 5. La “difettosità” dell'A.I. — 6. Responsabilità da prodotto e *product safety*: la necessaria distinzione tra *issues of permittance* ed *issues of liability*. — 7. Nesso di causalità, onere della prova e trasparenza. — 8. Riflessioni conclusive.

1. L'autonomia dell'A.I. ed i correlati problemi di riparto delle responsabilità

Lo specifico valore aggiunto dell'A.I. animata da algoritmi di *machine learning* risiede nella capacità della macchina di sviluppare diversi livelli di autonomia, modificandosi ed evolvendosi mediante un costante processo di *decision-making* ed auto-apprendimento, basato sia sul *training* ricevuto che sulla “esperienza” accumulata nel corso della propria funzionalità; dunque anche in un momento successivo all'uscita dell'A.I. dalla sfera di controllo del suo produttore. Tale elevato grado di autonomia, che consente alla macchina di eseguire *task* di varia natura anche senza l'apporto (ed il controllo) umano, ha persino indotto un numero sempre crescente di autori ad interrogarsi sull'opportunità di conferire alla macchina stessa uno *status* giuridico autonomo, diverso da quello di mera *res* (sul tema, si rinvia al contributo di Ugo Ruffolo, *La “personalità elettronica”*, in questo volume)¹.

A tali caratteristiche corrisponde il rischio, da più voci denunciato, di un più o meno elevato livello di opacità del funzionamento dell'A.I. e dei suoi processi decisionali; e, dunque, di “comportamenti emergenti” della macchina, ossia di sue condotte non sempre prevedibili in anticipo, né comprensibili o spiegabili *ex post*, e potenzialmente dannose, o comunque indesiderate (a tal riguardo, sia consentito rinviare al mio contributo *La governance dell'Intelligenza Artificiale: profili e prospettive di diritto dell'Unione Europea*, in questo volume)². Ne derivano,

¹ Il presente scritto anticipa parziali contenuti di un futuro più ampio lavoro monografico.

² BERTOLINI, *Robots as products: the case for a realistic analysis of robotic applications and liability rules*, in *Law, Innovation and Technology*, 2013, 214; BOSCARATO, *Who is responsible for a robot's actions?*

sul piano sia etico che giuridico, due fondamentali criticità. La prima: se il processo evolutivo e di *decision-making* dell'A.I. risulta difficilmente conoscibile o spiegabile anche dal suo progettista o produttore, potrebbe essere parimenti complesso “governare” la macchina ed eventualmente correggerne *output* errati o condotte indesiderate, prevenendo l'insorgere di eventi dannosi. La seconda: la mancanza di trasparenza del funzionamento dell'A.I. potrebbe porre più di un problema nel garantire tutela specialmente ai “soggetti deboli” (quali i consumatori) che entrano in contatto con la macchina e ne subiscono le “decisioni” e gli effetti, potenzialmente anche dannosi. Ne potrebbero conseguire significative difficoltà nel riparto delle responsabilità per danni cagionati da malfunzionamenti dell'A.I. o sue condotte imprevedute.

Tali difficoltà si complicano ulteriormente laddove si considerino i nuovi scenari dell'*Internet of Things* (IoT)³, nei quali la interazione e la reciproca integrazione di sistemi e *device* “intelligenti” e connessi crea una struttura a più livelli, al cui funzionamento contribuiscono molteplici *player*: non soltanto i produttori dei singoli *device* che costituiscono la rete dei prodotti interconnessi, ma anche i programmatori dei *software* che li animano (*embedded* o meno) e gli ideatori dei relativi algoritmi, nonché il gestore della rete che consente la loro connessione. In un tale contesto, la molteplicità dei soggetti coinvolti rischia di aumentare esponenzialmente la difficoltà di ripartire la responsabilità per eventuali malfunzionamenti (più o meno prevedibili) di uno o più prodotti facenti parte del *network*.

Ad oggi l'analisi della problematica è stata declinata, specialmente a livello UE, su due principali fronti. Da un lato, si è dato corso ad una riflessione sulle implicazioni etiche, specialmente per le imprese di settore, delle criticità sopra descritte; riflessione che è sfociata, allo stato, nella formulazione di Linee Guida Etiche per una A.I. “affidabile” da parte di un *Expert Group* nominato dalla

An initial examination of Italian law within a European perspective, in VAN DEN BERG, KLAMING (a cura di), *Technologies on the stand - Legal and ethical questions in neuroscience and robotics*, Nijmegen, 2011; KARNOW, *The application of traditional tort theory to embodied machine intelligence*, in *We Robot*, 1; MATTHIAS, *The responsibility gap: ascribing responsibility for the actions of learning automata*, in *Ethics and Information Technology*, 2004, 6, 175.

³ Con la locuzione “*Internet of Things*” si fa generalmente riferimento ad un complesso formato da *device* “intelligenti” ed interconnessi grazie alla rete. L'esempio più lampante è quello legato alle applicazioni di domotica evoluta (“*smart home*”), ma dinamiche omologhe saranno quelle con le quali sarà gestita, in prospettiva futura, la circolazione di *self-driving* o *driverless car* (il tema è trattato nel contributo di Ugo Ruffolo, *Intelligenza Artificiale ed automotive: veicoli self-driving e driverless e responsabilità*, in questo volume); senza dimenticare gli scenari dell'*Industrial Internet of Things* (tema esaminato, in questo volume, nel contributo di Alberto Vacchi). Sul tema, MARTIN-CASALS, *Causation and scope of liability in the Internet of Things (IoT)*, in LOHSE, SCHULZE, STAUDENMAYER (a cura di), *Liability for Artificial Intelligence and the Internet of Things*, Baden-Baden, 2019, 201.

Commissione⁴, le quali individuano una serie di parametri cui gli operatori del settore dovrebbero ispirare la creazione e l'impiego di forme di A.I. (sui temi dell'“etica dell'A.I.”, si rinvia ai saggi di Lorenzo D'Avack, Ugo Pagallo, Paolo Moro, Giovanni Sartor e Francesca Lagioia, in questo volume; per una panoramica degli interventi ad oggi posti in essere in sede UE sul tema dell'A.I., sia, inoltre, consentito il rinvio al mio *La governance dell'Intelligenza Artificiale: profili e prospettive del diritto dell'Unione Europea*, sempre in questo volume). Dall'altro lato, è stato avviato un processo di valutazione dell'idoneità delle attuali norme civilistiche per la responsabilizzazione, “a valle”, delle imprese di settore in relazione a danni cagionati da sistemi di A.I., con specifico riguardo al tema della responsabilità da prodotto difettoso⁵.

2. Le prospettive di revisione della Direttiva *Product Liability*

Le riflessioni circa la “indecifrabilità” dell'A.I. hanno condotto, a livello dell'Unione Europea, all'avvio di un processo di esame dell'attuale regime di responsabilità da prodotto difettoso, volto a valutare se le norme vigenti possano o meno continuare a garantire un adeguato livello di protezione dei consumatori con riguardo ai fenomeni dell'A.I. e dell'*Internet of Things*. È stato così avviato un processo di valutazione della Direttiva 85/374/CEE⁶, in relazione alla quale il gruppo di esperti a tal fine nominato ha, allo stato, rilevato la necessità di modifiche per adeguarla alle tecnologie della quarta rivoluzione industriale⁷. Necessità rilevata, da ultimo, anche dal Parlamento europeo⁸.

⁴ High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy Artificial Intelligence*, 8 aprile 2019.

⁵ A tal riguardo, la Commissione ha costituito un gruppo di esperti che opera in due diverse formazioni, di cui la prima è chiamata ad assistere nell'interpretazione ed aggiornamento della Direttiva 85/374/CEE in materia di *product liability* anche alla luce delle nuove tecnologie, e la seconda è chiamata a valutare se l'attuale regime di responsabilità risulti idoneo ad incentivare l'adozione di nuove tecnologie promuovendo la stabilità degli investimenti e la fiducia dei consumatori.

⁶ Direttiva 85/374/CEE del Consiglio, del 25 luglio 1985, relativa al “*ravvicinamento delle disposizioni legislative, regolamentari ed amministrative degli Stati Membri in materia di responsabilità per danno da prodotti difettosi*”.

⁷ Commissione Europea, *Working Document, Liability for emerging digital technologies*, 25 aprile 2018, COM(2018) 237 final. Sul tema si veda altresì il Report “*Liability for Artificial Intelligence and other emerging digital technologies*” licenziato il 21 novembre 2019 dall'*Expert Group on Liability and New Technologies* (formazione “Nuove Tecnologie”) nominato dalla Commissione Europea. Si veda altresì ALPA, *Programmi e iniziative dell'Unione Europea in materia di libera*

Pare, dunque, essere stata riconosciuta l'infondatezza dei rischi, paventati anche in sede unionale dai primi commentatori delle problematiche di responsabilità civile poste dall'avvento dell'A.I. (e sollevati, tra l'altro, dalla nota Risoluzione del Parlamento europeo del febbraio 2017 in tema di “*norme di diritto civile sulla robotica*”⁹), di vuoti di tutela (“*responsibility gap*”) in relazione a danni cagionati dall'A.I. e da forme di avanzata robotica “intelligente”.

È stato da più voci sostenuto, infatti, che il carattere di autonomia dell'A.I. e le criticità connesse alla mancanza di trasparenza ed esplicabilità rischierebbero di determinare, specialmente in materia di *product liability* (ma anche con riguardo al più generale ambito della responsabilità extracontrattuale), una interruzione del nesso causale tra condotta del produttore e danno cagionato dall'impiego del sistema *self-learning*; tanto, con specifico riguardo alla eventualità che la capacità di autoapprendimento della macchina la conduca a comportamenti dannosi impensati ed a scelte non prevedibili nemmeno dallo stesso suo produttore. Ci si chiedeva, pertanto, come potere addebitare al fabbricante di un prodotto *A.I.-powered* la responsabilità di un evento lesivo cagionato dal prodotto stesso in assenza di una effettiva possibilità, per il fabbricante medesimo, di prevedere che l'evento si sarebbe verificato. E si evidenziava, così, il rischio che la responsabilità per danni cagionati da condotte imprevedibili del prodotto intelligente potesse non trovare adeguata imputazione. Tali interrogativi venivano posti a giustificazione della teorizzata necessità di interventi normativi *ad hoc* e dell'introduzione di un nuovo regime di responsabilità specificamente concernente i “danni da A.I.”, diverso e parallelo rispetto a quello di *product liability*¹⁰.

I denunciati rischi di vuoti di tutela paiono, tuttavia, più percepiti che reali.

circolazione delle merci e responsabilità del produttore, in ALPA (a cura di), *La responsabilità del produttore*, Milano, 2019, 651.

⁸ Il riferimento è alla Risoluzione in tema di “*Automated decision-making processes*” adottata il 12 febbraio 2020. In tale documento, in particolare, si richiede alla Commissione di formulare proposte normative volte ad una revisione della Direttiva *Product Liability* prendendo atto delle inedite difficoltà nel riparto delle responsabilità per danni cagionati a consumatori a causa di processi di decisione automatizzata. Per un più approfondito esame del contenuto del documento richiamato, sia consentito rinviare al mio contributo, in questo volume, *La governance dell'Intelligenza Artificiale: profili e prospettive di diritto dell'Unione Europea*.

⁹ Parlamento europeo, Risoluzione del 16 febbraio 2017 recante “*raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica*”, 2015/2103(INL). Per un'analisi del testo della Risoluzione, sia consentito rinviare al mio *Robotica intelligente e responsabilità: profili e prospettive evolutive del quadro normativo europeo*, in RUFFOLO (a cura di), *Intelligenza Artificiale e responsabilità*, Milano, 2017, 77.

¹⁰ Parte della dottrina nordamericana nega la applicabilità delle norme in materia di *product liability* a sistemi di **Termine estratto capitolo** all'ambiente circostante e di

3.

Intelligenza Artificiale ed *automotive*: le responsabilità da veicoli *self-driving* e *driverless*

di Ugo Ruffolo

Sommario: 1. *Automated vehicle*, interrogativi etici e giuridici ed analisi economica del diritto. — 2. Veicoli *self-driving* e *driverless*: *quid novi*? — 3. Le responsabilità del produttore e la esigenza di un'allocazione non penalizzante dei costi da risarcimento. — 4. Il “difetto” del veicolo autonomo: vizi di costruzione e sicurezza del prodotto. — 5. La produzione di veicoli autonomi come “attività pericolosa”? — 6. Programmazione “etica” dell'*autonomous vehicle*? — 7. Le responsabilità dello “*human in command*” alla conduzione del veicolo autonomo, anche *ex art. 2054 c.c.* — 8. Responsabilità da “custodia” e possibile dislocazione remota dell'A.I. che governa il veicolo autonomo. — 9. “*No need to reinvent the wheel*”; una proposta di mediazione giuridica interpretativa e sistemica per governare le responsabilità da *highly automated vehicle*.

1. *Automated vehicle*, interrogativi etici e giuridici ed analisi economica del diritto

La regolamentazione dei “sistemi di trasporto intelligenti” — con particolare riguardo alle automobili *self-driving* e, in prospettiva, *driverless* — rappresenta uno dei più interessanti *case study* per l'analisi delle generali problematiche di mediazione giuridica poste dall'avvento dell'Intelligenza Artificiale (A.I.), e delle corrispondenti possibili soluzioni normative o interpretative, con specifico riguardo alla capacità di mediazione giuridica delle norme già esistenti. Tanto, in relazione sia agli aspetti regolatori di settore, di natura prevalentemente tecnica (requisiti per la omologazione e circolazione dei veicoli autonomi, *safety standard*, sicurezza stradale, gestione dei dati relativi al traffico, ...) ¹; sia — e tale sarà

¹ È ormai risalente la Direttiva 2010/40/UE, che tracciava un “*Quadro generale per la diffusione dei sistemi di trasporto intelligenti nel settore del trasporto stradale*”, seguita, in Italia, dall'emanazione, nel febbraio 2013, del Decreto del Ministero delle Infrastrutture e dei Trasporti concernente la “*diffusione dei sistemi di trasporto intelligenti*” e, poi, dal c.d. Decreto *Smart Roads* del 28 febbraio 2018, concernente “*Modalità attuative e strumenti operativi della sperimentazione su strada delle soluzioni Smart Roads e di guida connessa e automatica*”), attuativo della legge n. 205/2017. Norme, queste, di natura principalmente tecnica, che si limitano a delineare un *framework* volto, nelle intenzioni del legislatore, ad approntare a livello europeo le condizioni per la sperimentazione, l'implementazione e la futura diffusione dei nuovi mezzi di trasporto intelligenti. Dal 2010, poco (o nulla) è stato fatto sul piano normativo in termini sistemici, sia a livello UE, sia nell'ambito dei singoli ordinamenti domestici (con qualche isolata eccezione, quale la nota legge tedesca del 2017 sul tema della responsabilità da circolazione di veicoli autonomi). Il dibattito sul tema è stato ridestato dalla Risoluzione del Parlamento Europeo del febbraio 2017, recante “*raccoman-*

L'oggetto delle riflessioni che seguono — alle norme in materia di responsabilità per danni da utilizzo, detenzione, proprietà, fabbricazione e commercializzazione di *automated vehicle* (ed alle conseguenti esigenze di gestione e razionalizzazione dei costi d'impresa in relazione alle esposizioni a responsabilità, parallele alle esigenze di tutela di utenti e terzi, e connesse a quelle di nuove coperture assicurative).

Con riferimento a tale secondo ambito di indagine, ancora una volta il quesito riguarda, innanzitutto, quali siano le — eventuali — lacune da colmare con innovazioni normative e quali istanze di mediazione giuridica possano, invece, essere soddisfatte ricorrendo ad una interpretazione evolutiva delle norme vigenti; quali, in Italia, quelle del codice civile e del codice del consumo (per un'analisi più ampia della problematica circa la applicabilità, in via interpretativa, delle vigenti norme di responsabilità al settore dell'A.I., rinvio al mio contributo *Le responsabilità da Artificial Intelligence, algoritmo e smart product: per i fondamenti di un diritto dell'Intelligenza Artificiale self-learning*, in questo volume, nonché a quello *Verso una responsabilità da algoritmo?*, nel volume, da me curato, *Intelligenza Artificiale e responsabilità*, Milano, 2017).

Il dibattito dottrinale sul tema (ad oggi ancora relativamente esiguo) si è sinora concentrato soprattutto su specifiche tematiche di natura essenzialmente socio-etica o socio-economica, restando non sempre sufficientemente considerato come le specificità di settore possano essere “ricondotte a sistema”, ed effettivamente tradotte in strumenti di mediazione giuridica non episodica, ma sistematica. Troppo frequente è, infatti, il rischio che, a fronte delle peculiarità di nuove tecnologie, e della conseguente specificità di taluni troppo contingenti interrogativi “etici” di settore, si proclami la necessità di altrettanto specifiche e particolaristiche norme settoriali; con conseguenti rischi di inquinamento del sistema da “iperfetazione legislativa”.

Anche perché eventuali innovazioni normative settoriali rischierebbero di frapporre un serio ostacolo alle possibilità di coordinamento fra ordinamenti (nazionali e sovranazionali) e fra i differenti universi di *common law* e *civil law*. Laddove, per contro, soluzioni sistemiche unificanti consentirebbero di superare la divaricazione tra i due “pianeti” nella prospettiva di discipline uniformi (primaria esigenza dei mercati come della società civile globalizzata). Gli esiti di

dazioni alla Commissione concernenti norme di diritto civile sulla robotica” (seguita, poi, da una omologa Risoluzione parlamentare del febbraio 2019), la quale, *inter alia*, invocava la necessità di interventi normativi a livello unionale al fine di garantire lo sviluppo transfrontaliero di veicoli *self-driving*, e da una successiva Risoluzione del 15 gennaio 2019 in tema di “*autonomous driving in European transport*”, oltre che dalla Comunicazione “*Verso la mobilità automatizzata: una strategia dell'UE per la mobilità del futuro*”, pubblicata dalla Commissione Europea il 25 aprile 2018. Per un esame delle prospettive del diritto UE in materia si rinvia al contributo di Andrea Amidei *La governance dell'Intelligenza Artificiale: profili e prospettive di diritto dell'Unione Europea*, in questo volume.

tali operazioni di interpretazione sistematica relative ad ambiti di *civil law* potrebbero influenzare, anche in universi di *common law*, non solo le formulazioni di risposte normative ma altresì la evoluzione giurisprudenziale, e dunque il suo impatto di fatto normativo, veicolato dallo “*stare decisis*”.

La preminente attenzione ad oggi rivolta alle singole specificità di settore ed agli altrettanto specifici (e talora sopravvalutati e travisati) interrogativi etici rischia di far dimenticare due considerazioni che sono, invece, assorbenti nel dare risposta alle nuove istanze di mediazione giuridica nel settore degli *automated vehicle*.

La prima è che la diffusione dei veicoli autonomi attraverserà due fasi: una prima fase (già attuale), caratterizzata dalla circolazione di auto *self-driving* con guidatore vigile (e già ora circolano auto a guida sempre più automatizzata, ma con controllo umano, pur sempre decrescente); una seconda fase (per il cui avvento dovrà verosimilmente attendersi ancora qualche decennio), nella quale i veicoli saranno esclusivamente *driverless*, privi di comandi manuali, e la guida umana, troppo pericolosa, sarà vietata (o solo eccezionalmente consentita). E nella transizione verso la circolazione interamente automatizzata, il traffico sarà, per diversi lustri (se non decenni), misto e nelle strade conviveranno veicoli a guida umana e mezzi dotati di vari livelli di autonomia (v. Burns, Shulgan).

La seconda considerazione è che, nonostante l'eco mediatica delle notizie di — comparativamente — rari incidenti mortali imputabili alla circolazione sperimentale di auto *self-driving* o *driverless*, la introduzione di *automated vehicles* è suscettibile di condurre ad una considerevole diminuzione del numero dei sinistri (elevatissima nel futuro con circolazione di sole auto *driverless*, ma comunque consistente nel periodo “misto”). Tuttavia se, da un lato, saranno salvate non poche persone che sarebbero state altrimenti vittime della circolazione di mezzi tradizionali, si stima altresì, dall'altro, che il pur minor numero di incidenti con auto a guida autonoma lederà soggetti diversi, che non sempre sarebbero rimasti coinvolti in assenza dell'introduzione della nuova tecnologia. Un'analisi *cost-benefit* applicata sulla base delle due considerazioni che precedono ci induce subito ad una prima conclusione, anche in termini di analisi economica del diritto: la prospettiva di una elevatissima diminuzione degli incidenti stradali pone l'esigenza di evitare l'imposizione di *standard* settorialmente penalizzanti, suscettibili di ritardare l'avvento di nuove tecnologie il cui impatto, a livello di macro-fenomeno, sarà comunque benefico.

Si pensi, ad esempio, a quanto siano diffuse le tesi di chi pretende dalle “auto intelligenti” addirittura la capacità di valutare il numero di bambini a bordo del veicolo e su strada onde determinare quali soggetti siano più sacrificabili in caso di manovre d'emergenza (è l'abusato “dilemma del carrello”), anche sostenendo la necessità di interventi normativi in tal senso. I quali, tuttavia, ritarderebbero di non pochi anni la diffusione di tecnologie capaci di drasticamente ridurre gli incidenti stradali e di salvare un elevato numero di vite. È relativamente facile,

infatti, con la tecnologia già esistente, costruire auto con guida automatica ben più sicura di quella umana; ma pretendere da quella tecnologia la capacità (ancora irrealistica) di scelte etiche fondate sul censire l'età di ciascun essere umano, sulla strada e a bordo, significherebbe ritardarne la introduzione quasi *sine die*.

Occorre, dunque, pur tenendone debito conto, non sopravvalutare le particolaristiche questioni socio-etiche, chiedendosi, anche per esse, quanto la novità (e dunque la necessità di soluzioni con ricorso a nuove norme settoriali) sia reale e quanto (e quando) solo apparente, ed anzi spesso fonte di possibili antinomie ed incompatibilità con più generali esigenze etiche di settore. Altro è dare mediazione giuridica a scelte etiche non controvertibili, altro acriticamente privilegiare una sola faccia della medaglia, ignorando gli alternativi problemi etici che caratterizzano l'altra. Soprattutto, intenti prescrittivi quali quelli sopra descritti potrebbero, con ogni cautela, meglio tradursi in disciplina di natura "amministrativa", o comunque di formulazione dei requisiti tecnici necessari per omologare un veicolo od autorizzarne la circolazione, piuttosto che in precetti di responsabilità aquiliana settoriali.

E giungiamo anche ad un'altra conclusione preliminare, che richiama quanto detto sopra circa l'attenzione da prestare ai possibili effetti di innovazioni normative specificamente di settore, ed alle istanze di chi, sulla base di supposte esigenze tecnoetiche, chiede norme radicalmente nuove. Quantomeno nel medio-breve termine, nel periodo di circolazione mista fra le auto tradizionali e quelle *self-driving*, dalla guida autonoma non si potranno non esigere gli stessi criteri di comportamento esigibili dall'autista umano; e allora le forme di responsabilizzazione, quantomeno con riguardo alla responsabilità civile da circolazione, dovrebbero essere uniformi ed uguali per tutti, e non eccessivamente differenziate a seconda che la conduzione sia umana o automatizzata.

2. Veicoli *self-driving* e *driverless*: *quid novi*?

Aggiungiamo, poi, un'ulteriore riflessione. Il governo interamente umano (non solo della conduzione, ma anche) della trazione dei veicoli su ruote è fenomeno relativamente recente, sviluppatosi massivamente solo nello scorso secolo, e destinato a terminare nella prima metà di questo. Per millenni la trazione restò affidata alla intelligenza naturale: animali, esseri umani non *sui iuris* quali gli schiavi (liberati, negli Stati Uniti, solo alle soglie del Novecento). Fra qualche decennio essa sarà consegnata alla intelligenza artificiale. Da meno di due secoli, dunque, ed ancora per poco, al conducente umano è affidato il governo di una trazione non "intell...

Termine estratto capitolo breve — intervallo le sue