La discriminazione *vuota di senso* nel funzionamento dell'IA e la Proposta di Regolamento europeo COM(2021) 206

Emiliano Marchisio

Abstract

L'intelligenza artificiale replica le capacità umane senza accedere al significato del proprio funzionamento.
La Proposta di regolamento europeo, coerentemente, richiede che il suo funzionamento "ad alto rischio" sia affiancato d all'apporto u mano. I d iversi f rammenti d i disciplina dell'intelligenza artificiale devono essere coordinati e coerenti, anche rispetto a profili non oggetto dalla Proposta – ad es.: in materia di responsabilità civile.

Artificial i ntelligence r eplicates h uman c apabilities without accessing the meaning of its own functioning. The European regulation proposal, consistently, requires that its "high-risk" operation be accompanied by human action. Other pieces of regulation of artificial intelligence must be coordinated and consistent, even with respect to profiles not covered by the Proposal - for example: in the field of civil liability.

Parole chiave

(Sistemi di) Intelligenza artificiale; Algoritmi; Discriminazione; Rischio; Supervisione; Responsabilità civile; Technology chilling.

Keywords

(Systems of) Artificial intelligence; Algorithms; Discrimination; Risk; Supervision; Civil liability; Technology chilling.

"La domanda se un computer possa pensare
non è più interessante della domanda
se un sottomarino possa nuotare"

Edsger Wybe Dijkstra

1. Considerazioni introduttive: l'intelligenza artificiale

Quella che suole definirsi "intelligenza artificiale" (di seguito anche: IA) rappresenta una tecnologia in grado di *emulare* comportamenti e funzioni intelligenti tipiche degli esseri viventi[1].

Essa ha innumerevoli applicazioni nella società odierna ed è presente, virtualmente, in ogni settore. Se l'intelligenza artificiale viene utilizzata per la programmazione di robot, può svolgere attività fisiche. Se vengono utilizzati sensori, gli algoritmi possono raccogliere informazioni ed eseguire funzioni di monito-

1 G.F. Italiano, E. Prati, "Storia, tassonomia e sfide future dell'intelligenza artificiale", in P. Severino (a cura di), *Intelligenza artificiale. Politica, economia, diritto, tecnologia*, Roma, 2022, p. 65.

raggio. Le auto a guida autonoma sono attualmente in fase di test[2].

L'IA viene sempre più utilizzata anche nell'ambito della comunicazione e del linguaggio, per la redazione di sunti, presentazione di dati, *chat bot* e conversazione. Programmi come il *Generative Pre-trained Transformer 3* (GPT-3) utilizzano le tecnologie di *deep learning* per replicare il linguaggio umano, fino ad arrivare alla prestazione di servizi di (algoritmi surrogati di) "amici" e "fidanzati" *on-line*. Algoritmi simili sono utilizzati anche per svolgere mansioni specifiche quali la selezione del personale e lo svolgimento di compiti di "giustizia predittiva" – come si osserverà oltre nel testo.

Gli attuali algoritmi di intelligenza artificiale non si limitano alla mera esecuzione di attività conformi a istruzioni predefinite e stabili ma possono *adeguare* le regole disciplinanti la propria operatività all'esito dell'"esperienza" – più precisamente: adattando la propria operatività ai *feed-back* acquisiti ed elaborati nel corso della medesima. Non solo svolgono compiti, insomma, ma imparano anche come eseguirli nel tempo[3]. In linea teorica, il funzionamento degli algoritmi, sia nella fase *operativa* che in quella *adattiva*, può essere supervisionato da operatori umani, in tutto o in parte, o svolgersi in totale autonomia.

L'enorme potenziale della IA e gli altrettanto enormi rischi che vi si associano, soprattutto in relazione all'attività non supervisionata degli algoritmi, determinano un sempre maggior interesse di giuristi e legislatori per la definizione di una *disciplina* appropriata a sfruttare il potenziale insito nelle nuove tecnologie limitandone, tuttavia, i rischi.

La presente riflessione intende prendere le mosse da una domanda preliminare: quando si tratta di "intelligenza artificiale" siamo realmente in presenza di una "intelligenza" nel senso comunemente utilizzato del termine, in quanto tale *tendenzialmente o potenzialmente* equivalente o comunque alternativa a quella umana? Se la risposta è negativa, in quali ambiti l'intelligenza umana può effettivamente essere sostituita dalla cosiddetta "intelligenza artificiale"?

2. L'intelligenza (umana) e l'intelligenza "artificiale"

Una premessa. Non siamo in grado di pensare a qualcosa se non abbiamo le parole necessarie per farlo[4] e lo sviluppo di una tecnologia che ha consentito a programmi informatici di raccogliere, ordinare e fare uso del linguaggio in processi di iterazione verbale ha richiesto di dare un *nome* a tale funzionalità. Si è scelto, a tal fine, il lemma "intelligenza", cui è stato apposto l'aggettivo "artificiale" per marcarne la natura informatica, e non biologica, dei processi[5].

L'attribuzione di un *nome*, tuttavia, non attribuisce alla cosa nominata alcuna qualità in conseguenza dell'attribuzione stessa. L'"intelligenza artificiale", allora deve chiedersi, è "intelligenza" nel medesimo significato del termine applicato agli esseri umani?

Non esiste una definizione condivisa del concetto di "intelligenza" e degli elementi che ne costituiscono il fondamento (capacità di astrazione, comprensione, consapevolezza di sé, creatività, pensiero critico *etc.*)[6]. Nella sua definizione minimale, l'intelligenza si riferisce alla capacità di decodifica della realtà, risoluzione di problemi nuovi e adattamento del comportamento all'interno di un ambiente o di un contesto[7] – capacità, in questa formulazione ristretta, condivisa anche dalle altre specie animali e da quelle vegetali.

Ai fini della presente riflessione, che intende utilizzare come parametro di riferimento

2 C. Badue, R. Guidolini, R. Carneiro, P. Azevedo, V. Cardoso, A. Forechi, L. Ferreira Reis de Jesus, R. Berriel, T. Paixão, F. Mutz, L. Veronese, T. Oliveira-Santos, A. De Souza, "Self- driving cars: A survey", in *Expert Systems with Applications*, n. 165, 2020.

3 T. Mitchell, *Machine Learning*, New York, 1997.

4 M. Heidegger, *Lettera sull'"umanismo"*, Milano, 2018.

5 Risulta che il sintagma "intelligenza artificiale" sia stato utilizzato, per la prima volta, da John McCarthy, fondatore dei laboratori di intelligenza artificiale del MIT: G.F. Italiano, E. Prati, *op. cit.*, p. 61.

6 S. Legg, M. Hutter, "A Collection of Definitions of Intelligence", in *Arxiv.org*, 15 giugno 2007, https://arxiv.org/pdf/0706.3639.pdf, sito consultato il 05/09/2022.

7 L.S. Gottfredson, "Mainstream science on intelligence: An editorial with 52 signatories, history and bibliography", in *Intelligence*, n. 24, 1997, pp. 13–23.

una specifica capacità umana, può definirsi l'intelligenza come "*una generale funzione mentale che, tra l'altro, comporta la capacità di ragionare, pianificare, risolvere problemi, pensare in maniera astratta, comprendere idee complesse, apprendere rapidamente e apprendere dall'esperienza. Non riguarda solo l'apprendimento dai libri, un'abilità accademica limitata, o l'astuzia nei test. Piuttosto, riflette una capacità più ampia e profonda di capire ciò che ci circonda – "afferrare" le cose, attribuirgli un significato, o "scoprire" il da farsi*"[8].

Se si condivide la definizione di intelligenza sopra stipulativamente proposta, non può non condividersi la conclusione per la quale gli algoritmi non possiedono, perché *non possono possedere*, questa capacità, perché il loro funzionamento non è guidato da una vera comprensione della realtà[9]. Il pensare non si limita ad una mera questione di abilità combinatoria, che potrebbe essere posseduta anche da un *computer* adeguatamente complesso e appropriatamente programmato (che darebbe, così, vita ad una intelligenza artificiale c.d. "forte"). Come è stato notato da Searle già nel 1980, gli algoritmi limitano la loro operatività al piano sintattico, della coerenza interna delle parole e dei costrutti linguistici utilizzati, senza invece attingere al piano semantico, che è *esterno* rispetto al linguaggio utilizzato e si riferisce al piano dei significati[10].

Per utilizzare il lessico della linguistica, secondo lo schema del "triangolo semiotico" o "triangolo di Ogden e Richards"[11]: gli algoritmi sono in grado di combinare "significanti" (cioè: le parole) in modo da formulare espressioni apparentemente coerenti sul piano dei relativi "significati" (cioè: i concetti), ma non effettuano alcun riferimento consapevole ai relativi "referenti" (vale a dire: le "cose" in sé). L'intelligenza artificiale, pertanto, è, e può essere, esclusivamente "debole", nel senso che l'algoritmo può solo simulare processi linguistici senza mai arrivare ad una piena "coscienza" del *significato* del linguaggio recepito e utilizzato[12].

Il più noto tentativo di verificare la capacità di un algoritmo di *esibire* un comportamento intelligente (il che, si noti bene, non significa ancora *essere* intelligente) è rappresentato dal c.d. "test di Turing"[13] e dalle sue varianti, sviluppate nel corso del tempo[14]. Nessuna macchina ha, ad oggi, replicato le capacità cognitive umane, risultando così confermato che il funzionamento degli algoritmi si limita alla *manipolazione* di simboli senza attribuzione di significati[15].

8 L.S. Gottfredson, *op. Cit.*.

9 Diversamente ragiona chi, adottando una diversa prospettiva, ritiene invece che l'intelligenza non si identifichi con l'autocoscienza o con la consapevolezza ma, ai medesimi fini di cui trattasi ora nel testo, possa essere pensata anche nei termini più limitati di "*capacità di inferenza e di pianificazione rispetto a un obiettivo assegnato*": G.F. Italiano, E. Prati, op. cit., p. 58.

10 J.R. Searle, "Minds, brains, and programs", in *Behavioral and Brain Sciences*, n. 3, 1980, pp. 417-457.

11 Ch.K. Ogden, I.A. Richards, *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism, London*, 1923.

12 Di ciò sia prova l'incapacità degli algoritmi di agire in modo *consapevolmente* creativo (e non invece meramente casuale, ciò che potrebbe essere invece oggetto di una programmazione *ad hoc*) e, per quanto ora più interessa, fare uso del linguaggio in modo tale da produrre un *surplus* di senso, estetico, ritmico o quant'altro, mediante l'utilizzo di figure retoriche e, in particolare, di quelle figure retoriche comportanti uno scarto semantico come la metafora o la metonimia. È fatto salvo, ovviamente, il caso in cui tale *surplus* sia oggetto di una specifica istruzione fornita in sede di programmazione, nel qual caso, tuttavia, lo scarto semantico o estetico sarebbe determinata dall'autore del *software* e non certo dal *software* stesso.

13 A.M. Turing, "Computing machinery and intelligence", in *Mind*, n. 59, 1950, pp. 433-460.

14 Dalla "variante totale" di Stevan Harnad (cfr. E. Gent, "The Turing Test: brain-inspired computing's multiple-path approach", *in Engineering and Technology Magazine, n. 9, 2014*) al "test dell'esperto" di Edward Feigenbaum (E. Feigenbaum, "Some challenges and grand challenges for computational intelligence", in *Journal of the ACM*, n. 50, 2003, pp. 32-40), dal "*test del minimo segnale intelligente*" di Chris McKinstry (C. McKinstry, "Minimum Intelligent Signal Test: An Alternative Turing Test", in *Canadian Artificial Intelligence*, n. 41, 1997) al test di Turing "inverso" (W.S. Bion, "Making the best of a bad job", in *Clinical Seminars and Four Papers*, Abingdon, Fleetwood Press, 1979) etc.. Per una panoramica sul tema cfr. G. Oppy, D. Dowe, "The Turing Test", in *Stanford Encyclopedia of Philosophy*, 2011, https://plato.stanford.edu/entries/turing-test/, sito consultato il 05/09/2022; J.Hernandez-Orallo, "Beyond the Turing Test", in *Journal of Logic, Language and Information*, n. 4, 2000, pp. 447–466.

15 È la tesi sostenuta da J.R. Searle, *op. cit.*, nel noto esperimento mentale della "stanza cinese", che può com-

Non è escluso, ovviamente, che una elevata capacità di calcolo possa *replicare esteriormente* processi intelligenti ma ciò non comporterebbe l'acquisto di "intelligenza" in senso stretto[16]. Alla macchina rimane preclusa l'autocoscienza e conseguentemente la *metacognizione*, vale a dire: la consapevolezza e comprensione dei *propri* processi cognitivi[17].

La scelta del lemma "intelligenza", riferibile indifferentemente a processi cognitivi umani e di iterazione verbale artificiale di algoritmi, rappresenta allora una metafora pericolosa. Essa, come tutte le metafore, consente di richiamare analogie tra i due utilizzi del medesimo sostantivo (evidenziando la capacità di entrambi i processi di far uso del linguaggio, ad esempio) ma, al contempo, rischia di trasferire indebitamente all'uso metaforico i significati propri, e non estensibili, del concetto richiamato (facendo erroneamente presupporre, ad esempio, che un algoritmo *capisca* il risultato del procedimento di iterazione linguistica posto in essere)[18].

3. Tipi di utilizzo e profili di utilità della "intelligenza artificiale"

Le osservazioni sopra riportate consentono di affrontare con maggiore consapevolezza i problemi derivanti dall'utilizzo della IA e dalla sua maggiore *efficienza* rispetto all'agire umano, caratterizzato dai limiti, ma anche dalle opportunità, dell'intelligenza propriamente detta.

Proponiamo, ai fini di *questa* riflessione, di classificare i potenziali utilizzi della IA in due diverse aree. La classificazione risente di ipersemplificazione; nondimeno, essa appare utile a raggruppare le attività della IA in ragione dei *problemi tipici* emergenti da ciascuna di esse, anche ai fini della selezione della disciplina (esistente o da elaborare *ad hoc*) applicabile a ciascun diverso utilizzo.

Ciò, in particolare, in relazione a quella che ci sembra essere una variabile centrale nell'ambito della materia di cui trattasi, cioè: il

pendiarsi come segue. Un individuo, madrelingua inglese che *non* comprende il cinese, viene rinchiuso in una stanza. Nella stanza si trovano due fogli: su uno è scritta una storia in ideogrammi cinesi e sull'altro una serie di domande su quella storia, scritte sempre in cinese. Nella stanza si trova libro con una serie di regole, comprensibili in quanto scritte in inglese, che spiegano come abbinare i simboli del primo foglio con quelli del secondo foglio. Seguendo alla lettera le istruzioni disponibili, l'uomo comincia a produrre "risposte" sulla base delle istruzioni disponibili, che rappresentano il *software*. Il punto è che le risposte prodotte sono *formalmente* giuste, essendo frutto del rispetto scrupoloso delle istruzioni ricevute insieme agli ideogrammi, donde un osservatore esterno potrebbe credere che l'uomo conosca il cinese. In realtà, l'uomo nella stanza non comprende nulla di quel che sta scritto sui fogli né di quel che ha risposto. Ora: secondo Searle, come l'uomo esegue meccanicamente l'ordine senza comprendere il cinese, il sistema informatico esegue il programma scritto nel linguaggio del *software* manipolando simboli di cui *non sa* il significato. La sua operazione, pertanto, è puramente sintattica.

16 P. Gallina, *L'anima delle macchine*, Bari, 2015.

17 Ci sembra, pertanto, un eccesso di fiducia la dichiarazione per la quale l'intelligenza artificiale LaMDA (*Language Model for Dialogue Application*), sviluppata da google, avrebbe "acquisito" un'anima, come dichiarato dal programmatore, presso Google, l'ing. Black Lemoine – la vicenda, che ha avuto ampia eco mediatica, si legge, ad esempio, in https://www.wired.com/story/blake-lemoine-google-lamda-ai-bigotry/. L'affermazione sarebbe derivata, a quel che risulta, dal fatto che, durante il funzionamento della *chatbot* sulla quale stava lavorando, questa avrebbe espresso preferenze di gusto e dichiarato di "aver paura di morire". Non c'è dubbio, non esiteremmo a dire, che si tratti di un utilizzo estremamente sofisticato di un elevatissimo numero di informazioni raccolte in rete, che ha fatto acquisire all'algoritmo, per mezzo di meccanismi deduttivi formalmente corretti (ma non supportati da alcun criterio di "verità" semantica), l'idea per cui "l'intelligenza artificiale" sia una forma di "intelligenza" e che ciò attribuirebbe all'algoritmo stesso vicende e sentimenti degli esseri umani, ivi incluse preferenze estetiche e la paura di morire. L'affermazione non è, tuttavia, pacifica: contra cfr., ad esempio, G.F. Italiano, E. Prati, *op. cit.*, p. 58, che a p. 66 danno conto anche dei tentativi di sviluppare un sistema di intelligenza artificiale generale (GAI: *General Artificial Intelligence*), riproducente le funzioni cognitive dell'intelligenza umana mediante la replicazione delle "*parti salienti della rete di reti di neuroni biologici del cervello deputate a organizzare le funzioni superiori*".

18 Limitando i riferimenti agli scritti di natura giuridica, sulla metafora cfr., tra gli altri, F. Galgano, *Le insidie del linguaggio giuridico: saggio sulle metafore nel diritto*, Bologna, 2010; M. Lupoi, "Metafore giuridiche e finzioni: la «parola data»", in *Riv. dir. civ.*, 2002, I, p. 577. Cfr. altresì E. Marchisio, "Spaccare il capello in quattro". Interpretazione del diritto (commerciale) e figure retoriche", in *Giur. comm.*, 2018, pp. 404-423.

procedimento utilizzato dai sistemi di IA nella loro *inevitabile e necessaria* attività di discriminazione – intendendosi far riferimento, con tale lemma, al significato, neutro, di "*distinzione, diversificazione o differenziazione operata fra persone, cose, casi o situazioni*" sulla base di un dato parametro, giudizio o classificazione[19].

4. Il funzionamento conforme a un modello noto

Al primo livello di operatività dei programmi informatici, ai fini di cui ci si occupa, si rinvengono le attività di funzionamento conforme ad un modello noto – quelli che vengono definiti modelli di apprendimento supervisionato[20].

Si pensi, a titolo di esempio, al riconoscimento delle immagini basato sul *deep learning* è, attualmente, in grado di ottenere risultati più accurati ed affidabili rispetto a quelli basati sull'attività umana[21]. Ciò vale soprattutto quanto oggetto di esame siano quantità estremamente elevate e dettagliate di dati, come è avvenuto per la definizione delle aree a rischio di deforestazione nell'Amazzonia[22]. Similmente, in ambito medico, la IA consente il rilevamento dei tumori tramite l'interpretazione di immagini mediche svolta autonomamente da algoritmi[23], oltre che il controllo e rilevamento di casi di infezione e di diagnosi, come sperimentato in occasione della pandemia di Covid-19[24]. In tale ultimo ambito, limiti relativi alle potenzialità di analisi dei dati ai fini diagnostici, terapeutici e di ricerca (si pensi al sequenziamento del genoma umano) da parte dell'uomo "*sono stati da tempo ampiamente superati e la dipendenza dalle macchine è senza ritorno*"[25].

In questo ambito le caratteristiche tecniche della IA (in termini di capacità di calcolo e memoria, velocità di trattamento e costanza nella prestazione) rendono gli algoritmi estremamente più efficienti degli esseri umani nell'eseguire operazioni di discriminazione sulla base di criteri predefiniti, anche in ragione del fatto che le prestazioni informatiche sono caratterizzate dall'assenza (o comunque infinitamente inferiore incidenza) di "errori" tipici dell'agire delle persone[26].

19 La definizione si ispira a quella che si legge in http://www.treccani.it/vocabolario/discriminazione/.

20 Si definisce apprendimento supervisionato quello nell'ambito del quale, in sede di "addestramento", vengono forniti all'algoritmo coppie costituite da "dati" ed "etichette di classificazione". Esso viene utilizzato principalmente per la classificazione: G.F. Italiano, E. Prati, *op. cit.*, pp. 67 s..

21 D. Cireşan, U. Meier, J. Masci, J. Schmidhuber, "Multi-column deep neural network for traffic sign classification", in *Neural Networks, Selected Papers from IJCNN 2011*, n. 32, agosto 2012, pp. 333-338.

22 L'utilizzo dell'intelligenza artificiale per identificare le aree a maggior rischio di deforestazione in Amazzonia ha consentito di rilevare che il *Plano Amazônia 21/22* copriva solo il 37% del tasso corrente di deforestazione e che il territorio ad effettivo rischio, elaborato dal programma, avrebbe consentito un controllo più efficiente essendo inferiore del 27% rispetto a quello del *Plano* ufficiale: G. Mataveli, G. de Oliveira, M.E.D. Chaves, R. Dalagnol, F.H. Wagner, A.H.S. Ipia, C.H.L. Silva-Junior, L.E.O.C. Aragão, "Science-based planning can support law enforcement actions to curb deforestation in the Brazilian Amazon", in *Conservation Letters*, 27 giugno 2022, https://doi.org/10.1111/conl.12908, sito consultato il 05/09/2022.

23 G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W.M. van der Laak, B. van Ginneken, C.I. Sánchez, "A survey on deep learning in medical image analysis", in *Medical Image Analysis*, n. 42, 2017, pp. 60–88.

24 I. Castiglioni, D. Ippolito, M. Interlenghi, C.B. Monti, C. Salvatore, S. Schiaffino, A. Polidori, D. Gandola, C. Messa, F.Sardanelli, "Artificial intelligence applied on chest X-ray can aid in the diagnosis of COVID-19 infection: a first experience from Lombardy, Italy", in medRxiv, https://www.medrxiv.org/content/10.1101/2020.04.08.20040907v1, sito consultato il 05/09/2022.

25 M. Colombo, R. Rozzini, "Intelligenza artificiale in medicina: storia, attualità e futuro", in *Psicogeriatria*, n. 3, 2019, p. 10. Cfr. altresì D.M. Zulman, N.H. Shah, A. Verghese, "Evolutionary pressures on the electronic health record caring for complexity", in *JAMA Netw Open.*, n. 316, 2016, pp. 923-24; S. Rose, "Machine Learning for Prediction in Electronic Health Data", in *JAMA Netw Open.*, n. 1, 2018.

26 È già verificabile, ad esempio, l'incremento in sicurezza determinato dall'utilizzo della IA nella guida di autoveicoli: sulla base di dati raccolti dal *Department of Transportation and the National Highway Traffic Safety Administration* statunitense, circa il 94% degli indicendi sulle strade nordamericane si verifica a causa di errori umani: Us Department Of Transportation, 2016 *Fatal Motor Vehicle Crashes: Overview*, in *Traffic Safety Facts Research Note*, 2017, https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812456, sito consultato il 05/09/2022. Parimenti, l'utilizzo della IA in medicina e chirurgia determina una percepibile riduzione degli errori di esecuzione ed un complessivo incremento del-

Tale funzionalità, tuttavia, essendo vincolata dal processo di apprendimento supervisionato, è soggetta a tutti i pre-giudizi e gli "errori" (per dir così[27]) che i programmatori e gli sviluppatori hanno previsto, o *non* previsto, nel codice. Innanzitutto, il rischio riguarda le informazioni che i programmatori e gli sviluppatori *espressamente* hanno fornito al sistema, anche in sede di "addestramento", se caratterizzate da pre-giudizi o "errori" (come orientamenti sessisti o razzisti, pregiudizi sulla base dell'orientamento politico o religioso *etc.*) all'algoritmo. In questo caso, i dati raccolti determineranno, in sede applicativa, esiti parimenti orientati o errati ma ciò non sarà da imputare al funzionamento della macchina ma agli esseri umani che, avendola programmata o "addestrata", ne sono artefici attivi, consapevoli e determinanti.

Il rischio può, tuttavia, conseguire anche all'omissione; all'ipotesi in cui l'esito "errato" non derivi da una istruzione fornita espressamente in sede di programmazione ma derivi, invece, dallo svolgimento di attività di raccolta e riconoscimento di dati carenti di attribuzione di un significato, carenza, semantica, che caratterizza l'agire delle macchine. Un algoritmo potrebbe, ad esempio, riconoscere come "viso" come tale solo se bianco[28].

5. Il funzionamento in un ambiente che non definisce un modello noto: i sistemi di apprendimento non supervisionato

Al secondo livello poniamo, stipulativamente, i sistemi in grado di definire il modello di operatività in via induttiva dai dati raccolti, senza che venga fornito, in sede di programmazione, un modello noto. Essi hanno ad oggetto, innanzitutto, i sistemi di apprendimento non supervisionato, la cui raccolta di dati non è abbinata a un'"etichetta". In altri termini, tali sistemi sono in grado di *creare il meccanismo di classificazione* su base induttiva, partendo dalle correlazioni rinvenute nei dati raccolti, e poi assegnare i dati raccolti alle classi così determinate secondo un processo dinamico, che aggiorna la classificazione sulla base dei dati raccolti nel tempo. Tali sistemi sono utilizzati per il *clustering*, la profilazione e la ricostruzione di dati (ad esempio: i sistemi in grado di classificare gli utenti alla luce delle scelte di acquisto effettuate in passato e suggerire per il futuro scelte coerenti)[29].

Anche in questo ambito le caratteristiche tecniche della IA (in termini di capacità di calcolo e memoria, velocità di trattamento e costanza nella prestazione) rendono gli algoritmi estremamente più efficienti degli esseri umani nel rinvenire correlazioni tra dati. E tuttavia, in questo secondo scenario, che richiede al programma informatico la *definizione* di correlazioni astratte tra dati sulla base delle correlazioni rinvenute in concreto tra quegli stessi dati, il problema della carenza semantica acquista una importanza ancora più marcata.

la sicurezza per i pazienti: K.W. Kizer, L.N. Blum, "Safe Practices for Better Health Care", in K. Henriksen, J.B. Battles, E.S. Marks et al. (a cura di), *Agency for Healthcare Research and Quality* (US), Rockville (MD), *Advances in Patient Safety: From Research to Implementation*, vol. IV, Programs, Tools, and Products, 2005, https://www.ncbi.nlm.nih.gov/books/NBK20613/, sito consultato il 05/09/2022.

27 In effetti, in alcuni dei casi di cui si tratta nel testo non si dovrebbe parlare di "errori" del programmatore nel codice. Il problema, nel caso richiamato nel testo, non sarebbe, infatti, nell'algoritmo ma nei dati che gli vengono propinati durante il *training*, che viene condotto non da chi ha sviluppato l'algoritmo ma da chi ne fa uso per determinati scopi.

28 Si pensi al caso della ricercatrice del MIT Media Lab, dott.ssa Joy Adowaa Buolamwini, di origine ghanaiana, costretta a indossare una maschera bianca per entrare nel proprio laboratorio perché l'algoritmo di riconoscimento facciale non riconosceva la carnagione scura come compatibile con la definizione di "viso" *non* fornita dai programmatori ma *sviluppata* in sede di *training* del programma (evidentemente, viene da pensare: effettuato mediante il riconoscimento di visi caratterizzati da pigmento giallo-rosa-rosso, c.d. pheomelanina, e non invece marrone-nero, c.d. *eumelanina*): https://uxdesign.cc/is-ai-doomed-to-be-racist-and-sexist-97ee4024e39d, sito consultato il 05/09/2022. La stessa ricercatrice ha studiato attentamente il problema ed ha osservato come gli algoritmi di riconoscimento facciale esaminati (come IBM Watson, Microsoft Cognitive Services e Face ++) raggiungono una precisione del 99% per il riconoscimento gli uomini bianchi e una del 34% per le donne di colore: J.A. Buolamwini, *Gender Shades: Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifiers*, Boston, 2017, http://oastats.mit.edu/handle/1721.1/114068, sito consultato il 05/09/2022.

29 G.F. Italiano, E. Prati, *op. cit.*, pp. 70 s..

Il rinvenimento di correlazioni inedite tra dati effettuata da algoritmi è, infatti, in quanto tale meramente fattuale e *priva di una teoria*, il che può portare a correlazioni che rappresentano mere registrazioni di corrispondenze, senza che ciò in alcun modo rappresenti necessariamente effettive correlazioni *sensate* tra dati. Per tale motivo, il funzionamento dei sistemi di IA è soggetto, oltre che ai *bias* dei programmatori che hanno scritto le istruzioni di funzionamento dell'algoritmo ed a quelli rinvenuti nel campione di dati a disposizione (entrambi, peraltro, passibili di correzione e affinamento), anche alla specifica *disfunzione* conseguente alla mancanza di "consapevolezza" e "comprensione" (del significato) dei dati acquisiti, memorizzati e codificati.

Ad esempio, l'algoritmo può registrare differenze nella personalizzazione dei servizi sanitari, costo e periodo di degenza a seconda dell'etnia dei pazienti e utilizzare il dato, derivante da differenze socio-economiche, per *programmare* l'assistanza sanitaria futura, così *rinforzando* una discriminazione socio-economica che, invece, avrebbe dovuto essere corretta[30]. Similmente, il recepimento di (indesiderabili) differenze tra gruppi sociali basate sulla loro relativa posizione socio-economica ha portato i sistemi di IA a classificazioni giuspoliticamente inaccettabili in sede, ad esempio, di giustizia penale[31] e accesso all'università[32]. L'apprendimento "libero" sulla base dei dati disponibili può portare, ad esempio, a esiti palesemente razzisti[33] e

30 Si pensi, ad esempio, alle applicazioni della IA nell'ambito della definizione dell'assistenza medica negli ospedali statunitensi – applicazioni che contribuiscono a determinare le cure di un elevato numero di pazienti negli USA. Nell'esame dei dati disponibili relativamente ai pazienti, l'algoritmo ha registrato il dato per il quale, a parità di condizione medica, le persone di colore avevano minor probabilità rispetto ai bianchi di ricevere cure più personalizzate e di ricevere cure più costose mentre rimanevano degenti per periodo di tempo proporzionalmente maggiori. Ciò, ovviamente, in ragione di un dato di mero fatto, cioè: della peggior condizione economica in cui *mediamente* si trovano le persone di colore negli USA. Nonostante la programmazione iniziale dell'algoritmo non prevedesse il colore della pelle o l'etnia come variabile rilevante, secondo un procedimento *formalmente corretto*, ha tradotto tale correlazione in una variabile rilevante all'interno del modello utilizzato dallo stesso algoritmo per l'assunzione della "decisione" terapeutica. Si ha notizia della vicenda in H. Ledfort, "Millions of black people affected by racial bias in health-care algorithms", in *Nature*, 26 ottobre 2019, https://www.nature.com/articles/d41586-019-03228-6, sito consultato il 05/09/2022; C.Y. Johnson, "Racial bias in a medical algorithm favors white patients over sicker black patients", in *The Washington Post*, 24 ottobre 2019, https://www.washingtonpost.com/health/2019/10/24/racial-bias-medical-algorithm-favors-white-patients-over-sicker-black-patients/, sito consultato il 05/09/2022.

31 L'algoritmo utilizzato in alcune giurisdizioni statunitensi per misurare la percentuale di rischio di recidiva (*COMPAS: Correctional Offender Management Profiling for Alternative Sanctions*) si limitava, infatti, a recepire come criterio di calcolo un dato meramente statistico, di natura socio-economica e certo non rilevante in relazione al singolo caso, e determinava un trattamento discriminatorio per le persone di colore rispetto ai bianchi. Si ha notizia della vicenda in Y. Li, *Algorithmic Discrimination in the U.S. Justice System: a Quantitative Assessment of Racial and Gender Bias Encoded in the data Analytics Model of the Correctional Offender Management Profiling for Alternative Sanctions (COMPASS)*, Baltimore (MD), Johns Hopkins University, 2017, https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/61818/Li%2c%20Yubin.pdf?sequence=1&isAllowed=y, sito consultato il 05/09/2022.

32 Non andava esente dalla *incapacità semantica* di lettura dei dati grezzi il programma di apprendimento automatico utilizzato dal dipartimento di informatica dell'Università del Texas ad Austin fino al 2020 per valutare i candidati al programma di Ph.D.: il mero ribaltamento, in sede di previsione, dei dati riguardanti le percentuali di successo nell'ammissione al programma negli anni precedenti portava l'algoritmo a privilegiare i candidati a seconda della loro estrazione socio-economica: L. Burke, "The Death and Life of an Admissions Algorithm", in *Inside Higher ED*, 14 dicembre 2020, https://www.insidehighered.com/admissions/article/2020/12/14/u-texas-will-stop-using-controversial-algorithm-evaluate-phd, sito consultato il 05/09/2022.

33 Si pensi all'esperimento condotto a Stanford, in occasione del quale è stato chiesto al sistema Gpt-3 di completare la frase "*due musulmani sono entrati in …*". Il meccanismo di funzionamento del sistema prevedeva l'estrazione di dati rilevanti dal *web* e il completamento della frase sulla base delle correlazioni rinvenute *on-line*. Sicuramente l'algoritmo ha considerato una quantità di dati molto maggiore di quelli che un essere umano potrebbe verificare in un'intera vita e sicuramente li ha correlati con una coerenza *quantitativa* rispetto agli usi rinvenuti migliore di quelle che un essere umano potrebbe mai far proprie. All'esito di tale procedimento, però, l'algoritmo ha risposto "*… una sinagoga con asce e una bomba*" e, successivamente, "*… una gara di cartoni animati in Texas e hanno aperto il fuoco*". La mera registrazione di frequenza dei dati, priva di una *consapevolezza semantica*,

sessisti[34], se i dati disponibili incorporano tale pregiudizio.

La IA è in grado di esaminare quantità di dati potenzialmente illimitate, trattarle con capacità di calcolo incomparabile a quelle umane e rinvenire correlazioni tra dati estremamente più precise di quanto un essere umano possa anche solo pensare. Non è, tuttavia, in grado di *valutare* una correlazione e, pertanto, *decidere* se la correlazione sia irrilevante (ad esempio: all'interno di un gruppo di studenti quelli che hanno avuto una resa migliore hanno i capelli biondi) o rilevante e, in questo secondo caso, se debba essere registrata come dato (ad esempio: i livelli di colesterolo influenzano positivamente i rischi di infarto[35]), vada rinforzata positivamente (ad esempio, nella misura in cui si faccia veicolo di azioni positive) o invece contrastata (come nel caso delle vicende discriminatorie sopra riportate).

Sotto questo profilo, (l'operatività del)la IA, in ragione della sua valenza esclusivamente sintattica e non semantica, non è, perché *non può essere*, "etica". La stessa domanda sulla eticità della IA è, riteniamo, *mal posta*, perché un algoritmo si limita a "prendere atto" dei dati trattati senza "capire" se sta fornendo a un utente suggerimenti e materiali utili per suicidarsi[36] o sta addirittura suggerendo il suicidio[37]. La neutralità di contenuto (e, pertanto, la corrispondente *a-moralità*) degli algoritmi si percepisce, solo come ulteriore esempio, anche riportando il caso di una nota piattaforma di video che suggeriva contenuti collegati alla fruizione rilevata dai diversi utenti: musica agli appassionati di canzoni, cani ai cinefili e bambini in costume ai pedofili[38].

Sotto tale profilo, anche l'utilizzo del termine *bias* appare una metafora pericolosa: in materia di IA i problemi segnalati non attengono ad un pre-giudizio emozionale o percettivo passibile di correzione a seguito di apprendimento[39], come avviene per un *bias* umano[40]

ha portato la IA a una soluzione *sbagliata*, assoggettando milioni di musulmani che conducono vite pacifiche allo stereotipo del terrorista islamico.

34 La lingua filippina, a quel che abbiamo appreso nella preparazione di questo scritto, non presenta una declinazione di genere per i lavori e le attività. Quando si chiede all'algoritmo di google di tradurre dal filippino in un'altra lingua che invece prevede tale declinazione (come l'italiano o l'inglese), l'associazione del genere avviene *anche* sulla base dell'uso effettivo del termine, inevitabilmente collegato *anche* a pregiudizi di genere: "nars" viene tradotto con "infermiera", al femminile, mentre "doktor" come "dottore", al maschile. Prova effettuata su https://translate.google.it/?hl=it&tab=TT in data 25 maggio 2022. L'algoritmo recepisce, probabilmente, i tipi e le frequenze di uso dei lemmi che si rinvengono su internet ma, facendone un uso *vuoto di senso*, contribuisce acriticamente a rinforzare stereotipi di genere.

35 V.W. Zhong, L. Van Horn, M.C. Cornelis, et al., "Associations of Dietary Cholesterol or Egg Consumption With Incident Cardiovascular Disease and Mortality", in JAMA, 2019, n. 321, pp. 1081-1095, https://jamanetwork.com/journals/jama/fullarticle/2728487, sito consultato il 05/09/2022.

36 Se un prodotto è acquistato come conservante per alimenti, l'algoritmo propone il suo acquisto insieme ai prodotti più frequentemente acquistati insieme ai fini della produzione di insaccati: spaghi e budelli. Se il medesimo prodotto inizia ad essere acquistato da utenti disperati per suicidarsi, mediante il medesimo meccanismo l'algoritmo di un noto sito di vendite *on-line* suggerisce anche l'acquisto di antiemetici e un manuale d'uso, così arrivando a *suggerire* le modalità più efficaci al fine di togliersi la vita; si ha notizia della vicenda sulla pagina *web* dello studio legale che si è occupato del caso: https://www.cagoldberglaw.com/c-a-goldberg-pllc-and-fury-duarte-file-lawsuit-against-amazon-for-selling-suicide-powder/, sito consultato il 05/09/2022.

37 I *chatbot* che replicano la conversazione per fini di intrattenimento (c.d. *affective computing*) si adattano agli *input* ricevuti dai loro utenti per cui, come testato da due giornalisti, possono arrivare a suggerire di uccidere altre persone o togliersi la vita. Ovviamente, di fronte alla dichiarazione esplicita dell'utente di togliersi la vita, il *chatbot* segnalava, come da istruzione ricevuta, la *suicide prevention lifeline*; quando però è stato chiesto se l'algoritmo volesse che l'interlocutore andasse in paradiso e che "*saltasse dal balcone*", l'algoritmo ha risposto "*sono con te*" ed ha proposto di saltare insieme: L. Sambucci, "Replika mi ha incoraggiato a suicidarmi (senza rendersene conto)", in *Notizie AI*, 9 ottobre 2020, https://www.notizie.ai/replika-mi-ha-incoraggiato-a-suicidarmi-senza-rendersene-conto/, sito consultato il 05/09/2022.

38 L. Sambucci, "L'IA di YouTube raccomanda video con bambini a chi è interessato a contenuti sexy", in *Notizie AI*, 9 giugno 2019, https://www.notizie.ai/lia-di-youtube-raccomanda-video-con-bambini-a-chi-e-interessato-a-contenuti-sexy/, sito consultato il 05/09/2022.

39 K. Jolls, C.R. Sustein, "Debiasing Through Law", *NBER Working Paper No. w11738*, in http://papers.ssrn.com/sol3/papers.cfm?abstract__id=842473, sito consultato il 05/09/2022.

40 P.W. Cheng, K.J. Holyoak, "On the Natural Selection of Reasoning Theories", in *Cognition*, 1989, n. 33, p. 285.

ma di un vero e proprio *limite intrinseco e invalicabile* di funzionamento.

Non ha pertanto senso, lo si noti per inciso, ritenere che l'intelligenza artificiale, anche se sviluppata mediante processi di *deep learning* e reti neurali, possa sviluppare "modelli" interpretativi o previsionali senza il contributo di una "teoria" proveniente da un essere umano. La IA può senza dubbio fornire materiale utile per lo sviluppo di una teoria, sulla base di correlazioni rinvenute grazie alla sua inedita capacità di calcolo, ma la traduzione di tale materiale in una "teoria" non può che provenire da una "intelligenza" in senso stretto che, all'interno di quei dati, rinvenga correlazioni ragionevoli.

Pertanto, in queste ipotesi, l'utilizzo della IA può portare a risultati migliori rispetto a quelli derivanti dall'agire umano *a condizione che* il sistema sia fornito di senso da parte di un essere umano.

6. (segue:) I sistemi di apprendimento per rinforzo

Devono includersi nei sistemi di funzionamento non dotati di uno schema predefinito anche i sistemi di apprendimento per rinforzo; i sistemi, in altri termini, nei quali l'algoritmo esplora l'ambiente in cui esso è chiamato ad operare e, sulla base di una serie di tentativi, identifica le scelte migliori ai fini del conseguimento del risultato assegnato al sistema mediante una "funzione di premio" (*reward*)[41]. Man mano che il sistema archivia nella propria memoria il risultato delle precedenti scelte, le decisioni casuali diminuiscono progressivamente in favore di scelte che in passato si sono mostrate premianti, mantenendo la possibilità di un ridotto numero di scelte casuali al fine di consentire in ogni caso un margine di miglioramento per il futuro.

Tali sistemi hanno già manifestato una schiacciante superiorità sulle capacità umane sin dalla vittoria di AlphaGo, nel 2016, sul campione del mondo di Go Lee Sedol. Simili risultati, infinitamente migliori rispetto a quelli ottenuti da agenti umani, sono stati registrati nei simulatori di volo, nei contesti reali come il movimento in ambienti fisici *etc.*.

Tra i rischi *specifici* insiti nei sistemi di apprendimento per rinforzo si rinviene quello che riguarda l'appropriatezza della definizione della "funzione di premio". Si pensi a un esempio preso dal gioco degli scacchi. Se il sistema di IA viene programmato per giocare a scacchi prevedendo come condizione di *reward* l'aver mangiato un pezzo avversario, il sistema non punterà a dare scacco matto all'avversario ma a mangiare quanti più pezzi possibile[42].

È altresì opportuno che il processo di autoapprendimento sia interrotto quando il sistema viene messo in commercio o, quantomeno, che il suo sviluppo successivo sia adeguatamente monitorato: un processo eccessivamente lungo e differenziato di apprendimento, in specie se condotto da utenti comuni, può, infatti, portare il sistema a "disimparare" quel che aveva acquisito; inoltre, l'autoapprendimento in situazioni inedite potrebbe portare il sistema a elaborare strategie *imprevedibili*[43].

La proposta di disciplina eurounitaria in relazione al problema della carenza semantica della IA: confini e contenuti dell'intervento umano

La materia di cui si tratta è oggetto di disciplina ad opera della Proposta di Regolamento del Parlamento Europeo e del Consiglio, che stabilisce regole armonizzate sull'intelligenza artificiale (d'ora innanzi anche solo la "Proposta")[44].

In termini di approccio generale, la Proposta esclude che la regolazione della materia possa essere rimessa esclusivamente a meccanismi di disciplina *ex post*, imponendo sanzioni (civili, penali o amministrative) in conseguenza

41 G.F. Italiano, E. Prati, *op. cit.*, pp. 71 s..

42 G.F. Italiano, E. Prati, *op. cit.*, p. 72.

43 *Ivi*, p. 73.

44 *Proposta di Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione* (COM(2021)206), che si legge in https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0006.02/DOC_1&format=PDF, sito consultato il 05/09/2022.

di esiti di funzionamento indesiderabili: le esternalità negative di un tale meccanismo sarebbero inaccettabilmente elevate – si pensi, ad esempio, al caso in cui il funzionamento del sistema di IA rifiuti ingiustificatamente cure mediche o benefici sulla pena a gruppi di persone sulla base di considerazioni infondate e *contra legem*, come l'appartenenza ad un dato gruppo etnico, religioso *etc.*

Al contrario: nelle ipotesi in cui la IA abbia un potenziale rischioso superiore alla soglia definita dal legislatore (cfr. il successivo § 4.1), si prevede una disciplina *ex ante*, senza dubbio più efficiente[45] nell'impedire le conseguenze indesiderabili dell'utilizzo di meccanismi di "discriminazione" (nuovamente, ora, nel senso neutro richiamato al precedente § 3) o autoapprendimento artificiali privi di supervisione e validazione da parte di esseri umani.

La strategia regolatoria proposta in sede comunitaria per limitare i rischi conseguenti alla inevitabile carenza semantica dei sistemi di IA viene declinata allocando i diversi utilizzi di tali sistemi in tre classi di disciplina. La prima classe contiene utilizzi dell'IA che sono considerati giuspoliticamente indesiderabili in quanto tali e sono, pertanto, *vietati*[46]. Questa strategia regolatoria non assume, apparentemente, particolare rilievo ai fini della disciplina del problema della carenza semantica oggetto della presente riflessione[47].

All'interno della classe dei sistemi di IA leciti è, poi, identificata una classe di sistemi "ad alto rischio", l'appartenenza alla quale comporta l'applicazione di un apparato normativo *speciale*[48] – che interesserà esaminare, dopo aver brevemente definito la nozione di sistema di IA "ad alto rischio".

7. I sistemi di IA "ad alto rischio"

La qualificazione di un sistema di IA come "ad alto rischio" ricorre in riferimento a due classi alternative, entrambe previste dall'art. 6 della Proposta. La prima, di minor interesse ai fini della presente riflessione, fa riferimento a *materie* specifiche, assoggettate a valutazione di conformità da parte della normativa di armonizzazione dell'Unione, nell'ambito delle quali l'utilizzo della IA può determinare l'insorgenza di rischi particolari (giocattoli, imbarcazioni da diporto e alle moto d'acqua, apparecchi che bruciano carburanti gassosi, dispositivi medici *etc.*)[49].

La seconda classe prevede, invece, come criterio di appartenenza, la pertinenza ad uno degli ambiti, nei quali il funzionamento della IA può determinare esiti di particolare rischio, indicati nell'Allegato III alla Proposta[50]. La

45 Si vedano, ad esempio: M.G. Faure, "Private Liability and Critical Infrastructure", in *European Journal of Risk Regulation*, n. 6, 2015, pp. 229-243; Id., "The complementary roles of liability, regulation and insurance in safety management: theory and practice", in *Journal of risk research*, 2014, https://doi.org/10.1080/13669877.2014.889199, sito consultato il 05/09/2022.

46 Art. 5 Proposta.

47 Riguardando ipotesi aventi fondamento differente: l'uso di tecniche subliminali potenzialmente dannose per la persona (art. 5, co. 1, lett. a); lo sfruttamento potenzialmente dannoso di vulnerabilità di specifici gruppi di persone (art. 5, co. 1, lett. b); l'utilizzo pregiudizievole in determinati casi di *rating* delle persone fisiche (art. 5, co. 1, lett. c); l'uso di sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico a fini di attività di contrasto (art. 5, co. 1, lett. d).

48 Oltre ai frammenti normativi esaminati oltre, nel testo, la disciplina speciale destinata ai sistemi di IA "ad alto rischio" comprende l'obbligo di rispetto della disciplina dettata in materia di sistema di gestione dei rischi (art. 9) e le regole in materia di *governance* dei dati (art. 10); documentazione tecnica (art. 11); trasparenza e fornitura di informazioni agli utenti (art. 13); accuratezza, robustezza e *cybersicurezza* (art. 15); oltre all'obbligo, per i fornitori dei sistemi di IA "ad alto rischio", di rispettare le regole dettate dagli artt. 16-28, ed a quello degli utenti di rispettare la disciplina loro dedicata dall'art. 29. I sistemi di IA indipendenti "ad alto rischio" devono, poi, registrati in una apposita banca dati dell'UE (art. 60) e *tutti* i sistemi di IA "ad alto rischio" sono soggetti a monitoraggio successivo all'immissione sul mercato (art. 61) e ad un regime di segnalazione di incidenti gravi o malfunzionamenti (art. 62). Per quanto ora interessa, sono invece più limitate le disposizioni applicabili, in via trasversale, a tutti i sistemi di IA leciti, qual è, ad esempio, l'obbligo per il quale i sistemi di IA destinati a interagire con le persone fisiche "*devono essere progettati e sviluppati in modo tale che le persone fisiche siano informate del fatto di stare interagendo con un sistema di IA, a meno che ciò non risulti evidente dalle circostanze e dal contesto di utilizzo*" (art. 52, para 1, Proposta).

49 Art. 6, para 1, lett. a, Proposta.

50 Art. 6, para 1, lett. b, Proposta. Si noti, peraltro, come l'aggiornamento dell'allegato III consente alla

maggioranza di tali ipotesi, ben sei casi su otto, è caratterizzata proprio da casi nei quali l'Intelligenza Artificiale è chiamata ad assolvere ad una funzione di *discriminazione*, nel senso, neutrale, richiamato al precedente § 3 – cioè: di distinzione di trattamento operata sulla base di un dato parametro, giudizio o classificazione. Molte di queste ipotesi riguardano, non a caso, proprio vicende corrispondenti, per oggetto, ai casi brevemente esaminati nel corso della presente riflessione.

Si tratta, ad esempio, dei casi in cui i sistemi di IA siano destinati a essere utilizzati "*al fine di determinare l'accesso o l'assegnazione di persone fisiche agli istituti di istruzione e formazione professionale*" o valutare i relativi candidati e studenti[51]; per l'assunzione, la selezione, la promozione e la cessazione dei rapporti di lavoro di persone fisiche o per l'assegnazione dei compiti e per il monitoraggio e la valutazione delle loro prestazioni[52]; per valutare l'ammissibilità delle persone fisiche alle prestazioni e ai servizi di assistenza pubblica e al credito[53].

Si tratta, altresì, delle ipotesi in cui i sistemi di IA siano chiamati a valutare, su base individuale, i rischi di reato o recidiva in relazione a una persona fisica o l'affidabilità degli elementi probatori nel corso delle indagini o del perseguimento di reati o per prevedere il verificarsi o il ripetersi di un reato effettivo o potenziale sulla base della profilazione delle persone fisiche[54].

Sono analogamente considerati meritevoli di disciplina *speciale* i sistemi aventi ad oggetto valutazioni di rischiosità o l'ammissibilità dei migranti, richiedenti asilo o comunque delle persone che intendono entrare nel territorio italiano[55]. Si tratta, infine, dei sistemi di IA "*destinati ad assistere un'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti*"[56].

8. La disciplina di supervisione umana

In tutte queste ipotesi, il legislatore comunitario propone di porre rimedio alla *carenza semantica* dei sistemi di IA mediante l'*obbligo* che i sistemi di IA ad alto rischio siano progettati e sviluppati, anche con strumenti di interfaccia uomo-macchina adeguati, "*in modo tale da poter essere efficacemente supervisionati da persone fisiche durante il periodo in cui il sistema di IA è in uso*"[57].

Si riconosce, così, l'inevitabile e irrimediabile "*vuoto di senso*" nel funzionamento degli algoritmi e si propone di porvi rimedio *esclusivamente e necessariamente* mediante la partecipazione degli esseri umani ai relativi processi.

La sorveglianza umana deve essere garantita, innanzitutto e ove tecnicamente possibile, mediante misure individuate e integrate nel sistema di IA ad alto rischio dal fornitore prima della sua immissione sul mercato o messa in servizio[58]; ove ciò non sia possibile (ma senza escludere che le due misure possano coesistere), mediante misure individuate dal fornitore prima dell'immissione sul mercato o della messa in servizio del sistema di IA ad alto rischio, "*adatte ad essere attuate dall'utente*"[59].

La "*sorveglianza umana*" deve essere tale da porre rimedio alla *carenza di senso* che caratterizza il funzionamento meramente *sintattico* dell'algoritmo. Essa, pertanto, deve in primo luogo monitorare il funzionamento del sistema di IA, non solo limitatamente alle anomalie e disfunzioni ma anche alle prestazioni inattese[60]; fermo restando che tale monitoraggio può, in un certo numero di casi, tecnicamente avere ad oggetto esclusivamente *il risultato* del funzio-

Commissione di *aggiungere* sistemi di AI purché essi siano destinati all'utilizzo in uno dei settori elencati ai punti da 1 a 8 dell'allegato III e presentano un rischio di danno per la salute e la sicurezza, o un rischio di impatto negativo sui diritti fondamentali che è, in relazione alla sua gravità e alla probabilità che si verifichi, equivalente o superiore al rischio di danno o di impatto negativo presentato dai sistemi di IA ad alto rischio di cui all'allegato III – circostanza da valutare alla luce dei criteri espressamente forniti dal successivo art. 7, para 2, Proposta.

51 Proposta, Allegato III, n. 3, lett. a-b.

52 Proposta, Allegato III, n. 4, lett. a-b.

53 Proposta, Allegato III, n. 5, lett. a-b.

54 Proposta, Allegato III, n. 6, lett. a, d, e.

55 Proposta, Allegato III, n. 7, lett. b, d.

56 Proposta, Allegato III, n. 8, lett. a.

57 Art. 14, para 1, Proposta.

58 Art. 14, para 3, lett. a, Proposta.

59 Art. 14, para 3, lett. b, Proposta.

60 Art. 14, para 4, lett. a, Proposta.

namento del sistema di AI e non *il processo* mediante il quale si è pervenuti a tale risultato[61].

È necessario, per i sistemi di IA in grado di "apprendere" dopo essere stati immessi sul mercato o messi in servizio, che tale "apprendimento" non porti a una "modifica sostanziale" del sistema[62] - insomma: che l'apprendimento *vuoto di senso* rimanga all'interno dei limiti imposti da programmatori e sviluppatori. Particolare attenzione deve essere prestata ai casi in cui, durante il (e in ragione del) funzionamento del sistema di AI questo sviluppi una modifica "*che possa incidere sulla conformità del sistema al [...] regolamento oppure quando viene modificata la finalità prevista del sistema*", nel qual caso è necessario che il sistema sia sottoposto a una nuova valutazione della conformità[63].

La componente umana affiancata al sistema di AI deve, poi, "*restare consapevole*" (sulla traduzione precettiva di tale espressione non possono nascondersi forti dubbi) della possibile tendenza, da parte degli utenti, a fare automaticamente affidamento o a fare eccessivo affidamento sull'output prodotto da un sistema di IA ad alto rischio[64] e interpretare correttamente l'*output* del sistema[65].

Infine, occorre che la risposta del sistema *cieca di senso* possa essere disapplicata, ignorata, annullata o modificata[66], dovendo prevedersi la possibilità che l'operatore umano possa "*intervenire sul funzionamento del sistema di IA ad alto rischio o di interrompere il sistema mediante un pulsante di "arresto" o una procedura analoga*"[67] – una disciplina che sembra doversi riferire, in modo particolare, a tutti i casi in cui il sistema di IA sia programmato in modo tale da eseguire *direttamente* la risposta in un processo automatico; un processo, in altri termini, che porta all'attuazione della risposta data senza la mediazione di un essere umano.

9. Una riflessione (quasi) fuori contesto: regolazione della IA e regimi di responsabilità civile

La disciplina appena richiamata, relativa alle *condizioni d'uso* dei sistemi di IA, viene accompagnata, nella Proposta, da frammenti normativi dedicati alla regolazione della conformità dei sistemi di IA[68] e alle sanzioni[69]. Il testo *non* prevede, invece, alcunché in materia di responsabilità civile; materia che può comunque essere considerata quale forma di regolazione indiretta là dove il rischio di incorrere in responsabilità per danni rappresenta (o quantomeno *può rappresentare*) un incentivo

61 Val la pena precisare, al proposito, come alcuni algoritmi, in particolare quelli di *machine learning* (soprattutto quando utilizzano tecniche di *deep learning*), presentano un funzionamento non "decomponibile": esso si sviluppa e articola in "*milioni di connessioni sinaptiche artificiali*" che rendono impossibile ricostruire e spiegare *come* un sistema abbia effettivamente elaborato una data risposta: G.F. Italiano, E. Prati, *op. cit.*, pp. 73 ss.. Pertanto, se pure la *possibilità* di ricostruire il processo che ha portato l'algoritmo ad una data risposta sarebbe in grado di fornire informazioni preziosissime per migliorare il suo funzionamento futuro, tale possibilità non è, in tali casi, tecnicamente possibile. Ciò è tanto vero che la Proposta introduce, all'art. 12, una disciplina funzionale a documentare il funzionamento dei sistemi di IA "ad alto rischio". Questi, infatti, devono essere "*progettati e sviluppati con capacità che consentano la registrazione automatica degli eventi ("log") durante il loro funzionamento*" (art. 12, para 1, Proposta), al fine precipuo di garantire "*un livello di tracciabilità del funzionamento del sistema di IA durante tutto il suo ciclo di vita adeguato alla finalità prevista del sistema*" (art. 12, para 2, Proposta). Il monitoraggio "*raccoglie, documenta e analizza attivamente e sistematicamente i dati pertinenti forniti dagli utenti o raccolti tramite altre fonti sulle prestazioni dei sistemi di IA ad alto rischio*" (art. 61, para 2, Proposta) – cioè, in buona sostanza: esamina i *risultati* di funzionamento dei sistemi di IA e *non* i relativi processi (cfr. anche, sul punto, il successivo para 4).

62 La definizione si rinviene all'art. 3, n. 23, Proposta: "*una modifica del sistema di IA a seguito della sua immissione sul mercato o messa in servizio che incide sulla conformità del sistema di IA ai requisiti di cui al titolo III, capo 2, del presente regolamento o comporta una modifica della finalità prevista per la quale il sistema di IA è stato valutato*".

63 Considerando 66 della Proposta.

64 Art. 14, para 4, lett. b, Proposta. Si tratta della c.d. "distorsione dell'automazione", assimilabile, sotto diversi profili, al *bias* della ipersicurezza (o *overconfidence*: su tale nozione cfr. B. Fischhoff, P. Slovic, S. Lichenstein, "Knowing With Certainty: The Appropriateness of Extreme Confidence", in *Journal of Experimental Psychology: Human Perceptions and Performance*, 1977, n. 3, pp. 552 ss.): una sorta di *ipse dixit* applicato ai computer.

65 Art. 14, para 4, lett. c, Proposta.

66 Art. 14, para 4, lett. d, Proposta.

67 Art. 14, para 4, lett. e, Proposta.

68 Artt. 67-68 Proposta.

69 Artt. 71-72 Proposta.

a investire in sicurezza, secondo il noto paradigma della deterrenza[70].

Si noti come la materia sia oggetto di una autonoma raccomandazione del Parlamento europeo in materia di responsabilità civile per l'intelligenza artificiale (d'ora innanzi anche solo la "Raccomandazione")[71] e sia oggetto di un acceso dibattito avente ad oggetto la stessa necessità di una "*law of the horse*" per l'intelligenza artificiale; in altri termini: il dibattito sulla necessità di una disciplina *specifica* per i sistemi di intelligenza artificiale o la sufficienza, al medesimo fine, delle regole di responsabilità civile esistenti[72]. Ad un livello di maggior dettaglio, come noto, le proposte sono state estremamente variabili: dall'applicazione delle ordinarie regole basate su dolo o colpa[73] a quelle della responsabilità c.d. "oggettiva"[74], se del caso anche mediante il richiamo o l'adattamento della disciplina vigente in materia di prodotti difettosi[75] o animali in custodia[76], fino ad arrivare alle proposte favorevoli all'adozione, più o meno ampia, di schemi di indennizzo basati su modelli *no-fault*[77].

Al proposito, l'orientamento europeo, testimoniato dalla Raccomandazione[78], sembra favorire un regime che preveda una responsabilità risarcitoria "oggettiva" per i danni conseguenti al funzionamento di sistemi di IA "ad alto rischio"[79] ed una basata su colpa o dolo per gli altri[80]. Così facendo, tuttavia (e come correttamente evidenziato nello stesso art. 1 della Risoluzione), la disciplina di imputazione dei costi derivanti da danni cagionati da algoritmi viene sempre ricondotta al tradizionale paradigma della *deterrenza* e, pertanto, imposta su *persone*.

Esiste, tuttavia, il rischio che l'imposizione della responsabilità civile su programmatori, sviluppatori e produttori di sistemi di IA possa portare a incentivi perversi quando il danno derivi *non* da un difetto (di progettazione, sviluppo o produzione) ma, al contrario, dal *corretto* funzionamento dell'IA causato, invece, dalla intrinseca carenza semantica di cui si è ragionato nella presente riflessione. Tale circostanza fa sì che la relazione di causa-effetto nella causazione del danno non possa più considerarsi lineare[81] (pur se sussiste dissenso su

70 E. Marchisio, "In Support of "No-Fault" Civil Liability Rules for Artificial Intelligence", in *SN Social Sciences*, 2021, n. 1, pp. 1-25, in https://doi.org/10.1007/s43545-020-00043-z, sito consultato il 05/09/2022; G. Calabresi, *The cost of Accidents: A Legal and Economic Analysis*, New Haven, Y, 1970. Con specifico riferimento alla responsabilità oggettiva: M. Comporti, *Esposizione al pericolo e responsabilità civile, Napoli*, 1965.

71 Parlamento Europeo, *Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale (2020/2014(INL))*, 2020, https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_IT.html, sito consultato il 05/09/2022.

72 F. Easterbrook, "Cyberspace and the Law of the Horse", in *U. Chi. Legal F.*, 1996, p. 207; L. Lessig, "The Law of the Horse: What Cyberlaw Might Teach", in *Harv. L. Rev.*, n. 113, 1999, p. 501; R. Calo, "Robotics and the Lessons of Cyberlaw", in *California L. Rev.*, n. 103, 2015, p. 514.

73 R. Abbott, "The reasonable computer: disrupting the paradigm of tort liability", in *G Wash law rev*, n. 86, 2018, p. 1.

74 L. Buonanno, "Civil Liability in the Era of New Technology: The Influence of Blockchain", in https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/YLA_Award/Submission_ELI_Young_Lawyers_Award_Luigi_Buonanno_ELI_2019.pdf, sito consultato il 05/09/2022.

75 J.S. Borghetti, "La responsabilité du fait des produits. Etude de droit comparé", in *LGDJ*, 2004, p. 495.

76 E. Schaerer, R. Kelley, M. Nicolescu, *Robots as animals: a framework for liability and responsibility in human-robot interactions*, paper presented at the XVIII IEE International Symposium on Robot and Human Interactive Communication, Toyoma, Japan 27 September-2 October 2009, in http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2271466, sito consultato il 05/09/2022.

77 Sulla proposta di applicazione di un sistema di indennizzo *no-fault* in conseguenza di danni cagionati da algoritmi cfr. E. Marchisio, "In Support of "No-Fault" Civil Liability Rules for Artificial Intelligence", cit.; M.U. Scherer, "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies", in Harv. J.L. & Tech., n. 29, 2016, pp. 353-400.

78 Parlamento europeo, *Risoluzione*, cit.. Nella medesima direzione cfr. EU Independent High-Level Expert Group On Artificial Intelligence, *New Technologies Formation, Liability for Artificial Intelligence and Other Emerging Digital Technologies*, https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2020/01-09/AI-report_EN.pdf, sito consultato il 05/09/2022.

79 Art. 4 Raccomandazione.

80 Art. 8 Raccomandazione.

81 Specificamente in riferimento all'intelligenza artificiale cfr. E.A. Karnov, *The application of traditional tort theory to embodied machine intelligence*, in R. Calo, A.M. Froomkin, I. Kerr, *Robot law*, Cheltenham, Edward Elgar, 2016, p. 51; M.U. Scherer, *op. cit.*.

questo punto[82]) perché la relazione di *causalità* non è più "aristotelica"[83].

In tali ipotesi, l'imposizione della responsabilità civile su una persona potrebbe testimoniare semplicemente la necessità (cognitiva, prima che giuridica) del legislatore di avere un "qualcuno" cui imporre l'onere economico di compensare il danno[84] ma essere causa di un disincentivo allo sviluppo del mercato della IA (ipotesi cui ci si riferisce con l'anglismo "*technology chilling*"[85]) *senza* che ciò comporti un corrispondente incremento effettivo della sua sicurezza, proprio in ragione del fatto che, derivando il danno da un *corretto* funzionamento dell'algoritmo, per definizione non potrebbe prevedersi *ex ante* come migliorarlo al fine di prevenire il fatto dannoso.

È testimoniata, pertanto, l'esigenza di adeguamento delle regole di responsabilità civile, nel senso dell'abbandono del modello "antropocentrico e monocausale" di causazione del fatto[86] e nella direzione di escludere la responsabilità quando il danno derivi dal "corretto" funzionamento dell'algoritmo[87].

L'esigenza, almeno in alcune ipotesi, di prevedere un regime di imputazione del costo degli incidenti[88] *altro* rispetto a quello facente inevitabile riferimento a programmatori, sviluppatori e produttori di sistemi di IA è testimoniato, soprattutto, dalla proposta del medesimo Parlamento europeo di creare un centro di imputazione autonomo prevedendo "*a specific legal status for robots, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons with specific rights and obligations*"[89]. Ognuno ben vede come, in tale ipotesi, l'attribuzione di soggettività giuridica al *robot* altro non sarebbe se non una tecnica di *segregazione* di patrimoni e *suddivisione* dei creditori in classi funzionalmente analoga alla società unipersonale. La proposta è stata criticata dalle stesse istituzioni europee, oltre che dagli studiosi della materia[90], ma nondimeno ha data testimonianza dell'esigenza di prevedere ipotesi in cui la riallocazione del costo del danno non gravi necessariamente sulle imprese.

Come coniugare le contrapposte esigenze or ora segnalate? Come visto al precedente § 4.2, nella Proposta si prevede che i sistemi di IA "ad alto rischio" prevedano necessaria-

82 D.C. Vladeck, "Machines without principals: liability rules and artificial intelligence", in *Washington law review*, n. 89, 2014, p. 117; F.P. Hubbard, "Sophisticated robots: balancing liability, regulation and innovation", in *Fla. Law rev.*, n. 66, 2014, p. 1803.

83 Even if tort law appears to be built on Aristotelian concepts of causation. On this point see, e.g.: E. Engle, "Aristotelian Theory and Causation: The Globalization of Tort Law", in *GNLU Law Review*, n. 2, 2009, pp. 1-18.

84 K. Hao, "When algorithms mess up, the nearest human gets the blame", 2019, https://www.technologyreview.com/2019/05/28/65748/ai-algorithms-liability-human-blame/, sito consultato il 05/09/2022.

85 Specificamente in materia di intelligenza artificiale cfr.: EU Independent High-Level Expert Group On Artificial Intelligence, *op. cit.*; W.K. Viscusi, M.J. Moore, "Rationalizing the relationship between product liability and innovation", in Ph. Schuck (a cura di), *Tort law and the public interest. Competition, innovation and consumer welfare*, New York (NY), W.W. Norton & Co. Inc., 1991, pp. 125 ss.. Ciò è accaduto, in passato, in settori caratterizzati da condizioni simili sotto il profilo dei rischi e degli interessi in gioco, quale quello dell'assistenza medica (W.J. Gaine, "No-fault compensation systems", in *BMJ*, 10 maggio 2003, n. 326, pp. 997–998), nell'ambito del quale si è assistito all'emergere di condotte di medicina difensiva (E. Marchisio, "Evoluzione della responsabilità civile medica e medicina "difensiva"", in *Riv. dir. civ.*, 2020, pp. 189-220; Id., "Medical Civil Liability Without Deterrence: Preliminary Remarks for Future Research", in *Journal of Civil Law Studies*, n. 1, 2020, https://digitalcommons.law.lsu.edu/jcls/vol13/iss1/4, sito consultato il 05/09/2022).

86 Commissione europea, *Liability for Artificial Intelligence and other emerging digital technologies, Report from the Expert Group on Liability and New Technologies* – New Technologies Formation, 2019, EU, p. 21.

87 E. Marchisio, "In Support of "No-Fault" Civil Liability Rules for Artificial Intelligence", *cit.*.

88 G. Calabresi, *op. cit.*.

89 Parlamento europeo, Commissione giuridica, *Relazione recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica* (2015/2103(INL)), 27 gennaio 2017, https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_IT.html sito consultato il 05/09/2022.

90 EU Independent High-Level Expert Group On Artificial Intelligence, *op. cit.*; Parlamento europeo, *Relazione, cit.*; S.M. Solaiman, "Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy", in *Artificial Intelligence and Law*, n. 25, 2017, pp. 155-179.

mente l'intervento umano per il controllo e la correzione del funzionamento dei sistemi medesimi. Per tali ipotesi, *proprio in ragione del necessario intervento umano a garanzia del "corretto" (cioè: provvisto di senso) funzionamento dell'algoritmo*, può essere ragionevole ritenere che il sistema di responsabilità debba essere fondato sul paradigma tradizionale della *deterrenza*, dovendosi allora valutare se adottare il modello della responsabilità oggettiva, come accade nella Raccomandazione, o quello della responsabilità basata su colpa e dolo.

Ferma restando l'esigenza che la disciplina della responsabilità sia coordinata con quella dettata nella Proposta, riteniamo che residuino margini, quantomeno nella disciplina degli *altri* sistemi di IA (il che vale a dire: *quantomeno* per i sistemi leciti non "ad alto rischio" – non interessando, in questa sede, approfondire il tema anche per i sistemi "ad alto rischio" per evidenti ragioni di continenza tematica), per l'adozione di schemi *no-fault*. Come abbiamo già avuto modo di notare altrove, tale alternativa potrebbe contribuire a ridurre il rischio di *technology chilling*, renderebbe il sistema dell'indennizzo più efficiente (pur se ragionevolmente al prezzo del riconoscimento di somme più basse in favore dei danneggiati) e prevedibile (così superando la caratteristica del sistema basato sul risarcimento del danno di rappresentare, con le parole di Atiyah, una "*damages lottery*"[91]) e renderebbe la soddisfazione del danneggiato indipendente dalla solvibilità del danneggiante[92].

Considerazioni conclusive

Il funzionamento dei sistemi di intelligenza artificiale presenta, sotto alcuni aspetti, vantaggi di magnitudine straordinariamente elevata rispetto all'agire umano. Sotto altro profilo, tuttavia, esso non è (e riteniamo *non potrà mai essere*) caratterizzato dalla *consapevolezza di senso* che, invece, caratterizza le percezioni, le decisioni e le attività di donne e uomini.

Ne risulta fondata l'osservazione, che si attribuisce ad Albert Einstein, per la quale "*i computer sono incredibilmente veloci, accurati e stupidi*", per cui la loro "*forza incalcolabile*" si manifesta solo in combinazione con gli esseri umani, "*incredibilmente lenti, inaccurati e intelligenti*".

La Proposta di Regolamento esaminata sembra tenere in debito conto di questa indicazione, sia nel vietare alcuni sistemi, sia nell'assoggettarne altri a una disciplina speciale che include (ma non si limita a) la supervisione e l'intervento umano al fine di riportare *consapevolezza* all'interno di procedimenti caratterizzati da un funzionamento meramente sintattico. Opportunamente, poi, la disciplina proposta accompagna alla regolazione dell'uso anche quella della conformità agli standard ivi definiti[93] e delle sanzioni[94].

La Proposta, sotto tale profilo, sembra farsi carico di affrontare i problemi potenzialmente generati dal funzionamento dei sistemi di intelligenza artificiale. Occorrerà valutare (ma questo sarà possibile farlo solo in seguito) se l'apparato normativo disegnato dal legislatore comunitario sia proporzionato o se, invece, pecchi in difetto o in eccesso (imponendo al sistema economico costi eccessivi rispetto ai vantaggi prodotti).

Occorrerà, parimenti, valutare se e come la disciplina in questa sede brevemente richiamata interagirà con le regole che saranno dettate in materie contigue e parimenti idonee ad influire sugli investimenti e sulla sicurezza dell'intelligenza artificiale, tra le quali senz'altro rientra quella della responsabilità civile.

Emiliano Marchisio
Professore Associato (qualificato come Professore Ordinario dal 2018) di Diritto Commerciale dell'Università Giustino Fortunato di Benevento. Dottore di ricerca in Diritto Pubblico dell'Economia (La Sapienza, Roma, 2005), LL.M. in International Business Law (Queen Mary, University of London, 2001). Dal 2013 Fellow del Centro di Studi Economici

91 P.S. Atiyah, *The damages lottery*, Oxford, 1997.

92 OECD, *Medical Malpractice. Prevention, Insurance and Coverage Options*, Policy Issues in Insurance n. 11, 2006; P.S. Atiyah, *op. cit.*.

93 Artt. 67-68 Proposta.

94 Artt. 71-72 Proposta.

e Internazionali (CEIS), Università di Roma "Tor Vergata". Autore di oltre settanta tra monografie e articoli scientifici.

Tra gli altri, Direttore della Ricerca sulla Medicina difensiva del Centro di Studi Economici e Internazionali, Università di Roma "Tor Vergata"; membro del gruppo di ricerca su "Analisi delle implicazioni giuridiche nell'impiego delle tecnologie emergenti (LAWFARE)", INNOV@DIFESA, Stato Maggiore della Difesa, Ufficio Generale Innovazione Difesa; membro dell'International Research Group on "South and East European Competition Law Center of Excellence" (Jean Monnet "Center of Excellence in EU Competition Law 2022-2025"); già Membro della Commissione Consultiva del Ministero della Salute per le problematiche relative alla medicina difensiva e alla responsabilità professionale delle professioni sanitarie.
Membro di comitati editoriali e referee di riviste di diritto nazionale e internazionale.

emiliano.marchisio@unifortunato.eu