JUS CIVILE

M. KEARNS, A. ROTH, *The Ethical Algorithm. The Science of Socially Aware Algorithm Design*, Oxford, 2020

L'algoritmo "etico" (a proposito di un libro recente)

È possibile imporre regole "etiche" all'intelligenza artificiale? O i tentativi di elaborare una disciplina giuridica capace di incanalarne gli sviluppi tumultuosi – oltre a risultare velleitari per insuperabili ragioni tecniche – sono forieri di inaccettabili vincoli alla ricerca scientifica, tanto da rendere incolmabile il *gap* fra l'Europa e le grandi potenze che dominano l'infosfera?

Il dibattito si è acceso a seguito della consultazione avviata dalla Commissione europea nel febbraio 2020 tra i principali *stakeholders* [1], con la pubblicazione del *White paper* sull'intelligenza artificiale [2], a cui ha fatto seguito la presentazione della Proposta di regolamento del Parlamento europeo e del Consiglio [3], contenente regole armonizzate sull'intelligenza artificiale. La sintesi delle varie posizioni è confluita nella citata proposta, che sembra incentrare il suo nucleo valutativo principalmente su un approccio basato sulla gestione del rischio [4]: i primi commentatori, tra critiche di impostazione e plausi su alcuni aspetti (come la *sandbox*) [5], sono concordi nel richiamare il *vulnus* maggiormente arduo da riempire, ovvero la responsabilità.

Fin dalla fase di consultazione, tuttavia, l'attenzione [6] si è concentrata sul profilo "eti-

---

[1] Conclusasi nel giungo 2020 e i cui risultati sono consultabili su https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence.

[2] https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

[3] "*Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts*". COM(2021) 206 final.

[4] V. considerando n. 14 della Proposta.

[5] Tra cui v. L. FLORIDI, *The European Legislation on AI: A Brief Analysis of its Philosophical Approach*, (June 1, 2021). Available at SSRN: https://ssrn.com/abstract=3873273; S. RANCHORDAS, *Experimental Regulations for AI: Sandboxes for Morals and Mores*, (May 4, 2021), University of Groningen Faculty of Law Research Paper No. 7/2021, Available at SSRN: https://ssrn.com/abstract=3839744; G. PROIETTI, *Intelligenza artificiale: una prima analisi della proposta di regolamento europeo*, Maggio 2021, in dirittobancario.it.

[6] In particolare da parte di tecnici ed imprese: cfr. E. KAZIM, A. KOSHIYAMA, *Lack of Vision: A Com-*

co" dell'intelligenza artificiale, sulla scia di quanto già ampiamente discusso tra gli accademici statunitensi, molti dei quali si sono espressi in senso favorevole alla creazione di regole chiare in grado di veicolare la ricerca entro binari "virtuosi" [7]. A sostegno di tale orientamento porta ora nuovi argomenti il recente saggio di due scienziati esperti di informatica teorica, il cui tentativo è spiegare «*how to design socially better algorithms*» [8].

In realtà, la propensione a fondare una scienza della ragion pratica dell'intelligenza artificiale sembra essere un cruccio particolarmente radicato proprio tra le fila degli studiosi delle "scienze dure" che si occupano di AI [9] con l'obiettivo, più o meno dichiarato, di consegnare all'umanità algoritmi più equi o "giusti".

Il libro dei due autori statunitensi Michael Kearns e Aaron Roth offre una riflessione meta-informatica al problema della "giustizia" dell'algoritmo, lasciando in penombra il tema della responsabilità e offrendo al lettore possibili soluzioni (meta-tecniche) per impegnare la ricerca scientifica in una specie di "ortopedia" dell'intelligenza artificiale che possa essere collettivamente edificante [10].

---

ment on the EU's White Paper on Artificial Intelligence, (March 20, 2020). Available at SSRN: https://ssrn.com/abstract=3558279; S. VON STRUENSEE, *Analyzing Dilemmas Posed by Artificial Intelligence and 4IR Technologies Post COVID-19, Requires using all Available Models, Including the Existing International Human Rights Framework and Principles of AI Ethics* (June 25, 2021). Available at SSRN: https://ssrn.com/abstract=3874279.

[7] V., per una sintesi del dibattito, J. FJELD, N. ACHTEN, H. HILLIGOSS, A. NAGY, M. SRIKUMAR, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI* (January 15, 2020). Berkman Klein Center Research Publication No. 2020-1, Available at SSRN: https://ssrn.com/abstract=3518482.

[8] M. KEARNS, A. ROTH, *The Ethical Algorithm. The Science of Socially Aware Algorithm Design*, Oxford, 2020, p. 101.

[9] E lo dimostra il copioso numero di saggi (anche a carattere divulgativo), elaborati da diversi esperti del settore, prevalentemente di origine anglo-americana: C. O'NEIL, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York, 2016; S. WACHTER-BOETTCHER, *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech,* New York-London, 2017; V. EUBANKS, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, New York, 2018; S.U. NOBLE, *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York, 2018; M. BROUSSARD, *Artificial Unintelligence: How Computers Misunderstand the World*, Cambridge, Massachusetts, 2018; H. FRY, *Hello World: Being Human in the Age of Algorithms*, New York-London, 2018 (trad. it. *Hello world. Essere umani nell'era delle macchine*, Bollati Boringhieri, 2019); G.F. MARCUS, E. DAVIS, *Rebooting AI: Building Artificial Intelligence We Can Trust,* New York, 2019; R. BENJAMIN, *Race After Technology: Abolitionist Tools for the New Jim Code,* Cambridge, 2019. Si segnala, nella letteratura italiana, il recente saggio di R. CUCCHIARA, *L'intelligenza non è artificiale. La rivoluzione tecnologica che sta già cambiando il mondo*, Mondadori, 2021.

[10] E v. M. KEARNS, A. ROTH, *The Ethical Algorithm*, cit., p. 94 ss. per l'applicazione dell'*algorithmic game theory* o p. 137 ss. sul *preventing p-hacking*.

Probabilmente, l'insegnamento più rilevante e, per certi versi, più innovativo, almeno nella prospettiva del giurista, può essere ravvisato nella strategia suggerita per arginare la "dis-umanità" dell'intelligenza artificiale: valorizzare il ruolo del *designer*. Infatti, a prescindere dalla caratura diversamente autonoma dell'algoritmo (sino ad arrivare al c.d. *machine learning* non supervisionato), il problema principale che gli scienziati si pongono è racchiuso indissolubilmente nell'ultimo "contatto" umano con l'artificiale; e una attenzione al fattore "umano", così fortemente invocata da parte di coloro che "creano" l'AI, è decisamente il valore aggiunto che deve indurre i giuristi a riflettere.

Come è stato messo in luce, gli algoritmi (e in particolare i modelli costruiti attraverso il *machine learning*) sono diversi, sia perché esercitano una significativa dose di autonomia nel prendere decisioni senza l'intervento umano, sia perché sono spesso così complessi e opachi che nemmeno i loro *designer* sono in grado di anticiparne il comportamento in molte situazioni [11]: quindi, il primo profilo che si snoda nella riflessione è quello dell'*agency,* mutuando il rapporto non tra due soggetti, ma tra un soggetto (il *designer*, appunto) e l'oggetto (l'algoritmo), avendo oramai superato l'epoca dell'*imitation game* [12].

Nel processo di mediazione, allora, ci si interroga sulle modalità attraverso le quali ciò contribuisce a veicolare i modelli, attraverso il "come" sia possibile progettare degli algoritmi maggiormente apprezzabili sul piano sociale: l'idea di fondo che sembra ispirare l'intera opera appare essere condensata nella capacità di orientare la "scelta" non dell'algoritmo bensì del *designer* [13].

In realtà, in questo frangente, non solo sono poste le basi per l'edificazione di un'"etica dell'algoritmo" – dove, accanto alla definizione dei concetti che maggiormente sono stati oggetto di attenzione nell'ultima decade (come la *privacy* [14] o la *fairness* [15]) e alla loro "implementazione" algoritmica, si stagliano nuovi criteri, quali la precisione (*accuracy*) e l'etica [16] – ma viene consegnata, probabilmente, l'idea più importante per

---

[11] M. KEARNS, A. ROTH, *The Ethical Algorithm*, cit., p. 7.

[12] R. CUCCHIARA, *L'intelligenza non è artificiale*, cit., p. 117 ss.

[13] Cfr. L. FLORIDI, *The Ethics of Information*, Oxford, 2013; ID, *The Logic of Information: A Theory of Philosophy as Conceptual Design*, Oxford, 2019.

[14] M. KEARNS, A. ROTH, *The Ethical Algorithm*, cit., p. 50: «*At its core, differential privacy is meant to protect the secrets held in individual data records while allowing the computation of aggregate statistics. But some secrets are embedded in the records of many people. Differential privacy does not protect these*».

[15] *Ivi*, p. 69 ss. per la *statistical parity*; p. 72 ss. per i correttivi alla *statistical parity* quali l'*equality of false negatives/positives* e (p. 78 ss.) il *Pareto frontier*. Per un'indagine di stampo giuridico, v. D.L. BURK, *Algorithmic Fair Use*, in *The University of Chicago Law Review*, Vol. 86, No. 2, 2019, p. 283 ss.

[16] *Ivi*, p. 78 ss.

cui non occorre elaborare istituti più o meno analoghi alla soggettività giuridica [17], ma bisogna concentrare gli sforzi attorno alla catena del valore dell'intelligenza artificiale: nel sinallagma algoritmo-software-applicazione, vi è evidentemente l'intervento dell'uomo, sia nell'implementazione dei dati [18] che nel *design* dell'intero circuito [19].

Questi ultimi canoni, a loro volta, producono necessariamente una serie di costi, in termini di accuratezza e precisione del modello stesso e in termini di quantificazione del *trade-off*[20]: dunque, la soluzione offerta è quella di rimettere la decisione, ancora una volta, nelle mani del *designer*, il quale diviene strumento di mediazione delle decisioni e dei sentimenti espressi dagli *stakeholders* [21]. A tal proposito, vengono riecheggiate proposte sulla falsariga della rivisitazione critica [22] della *Doctrine of the Double Effect* [23], dove, sulla distinzione tra fare e permettere, e tra intenzione diretta e obliqua, viene introdotta, tramite esperimenti mentali e dilemmi morali, la differenza tra doveri positivi e negativi e tra doveri in senso stretto e atti di carità, dando conto dei casi nei quali il dovere negativo sia preponderante su quello positivo e fornisca utili indizi per comprendere quale sia l'azione eticamente corretta da intraprendere [24]: insomma, tale prospettiva aiuta

---

[17] Per tutti, v. G. TEUBNER, *Ibridi ed attanti. Attori collettivi ed enti non umani nella società e nel diritto* (trad. it.), Milano-Udine, 2015; ID, *Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten*, in *AcP*, 218 (2018), p 155 ss., donde la soluzione per il riconoscimento agli agenti *software* lo *status* di attori parzialmente provvisti di capacità giuridica («*Für das Autonomierisiko ist es eine adäquate Antwort, den Softwareagenten den Status als teilrechtsfähige Akteure zuzuerkennen*», p. 204).

[18] Ed anche qui è possibile parla di etica dei dati: sul punto, cfr. L. FLORIDI, M. TADDEO, *What is data ethics?*, in *Phil. Trans. R. Soc. A 374*, p. 1 ss.

[19] R. CUCCHIARA, *L'intelligenza non è artificiale*, cit., p. 109 ss.

[20] M. KEARNS, A. ROTH, *The Ethical Algorithm*, cit., p. 192: «*One central lesson of this book is that additional constraints—like those imposed to correct ethical failures—won't come for free. There will always be trade-offs that we need to manage. Once we can precisely specify what we mean by "privacy" or "fairness," achieving these goals necessarily requires giving up on something else that we value—for example, raw predictive accuracy. It is the goal of algorithmic research not only to identify these constraints and embed them into our algorithms but also to* (p. 193) *quantify the extent of these trade-offs and to design algorithms that make them as mild as possible*».

[21] *Ivi*, p. 194: «*when we use algorithms not just to make predictions but to make decisions, they are changing the world in which they operate, and we need to take into account such dynamic effects in order to talk sensibly about something like "fairness"*».

[22] Ad opera di P.R. FOOT, *Natural Goodness*, Oxford, 2001; EAD, *Moral Dilemmas and Other Topics in Moral Philosophy*, Oxford, 2002; EAD, *Virtues and Vices and Other Essays in Moral Philosophy*, Oxford, 2002.

[23] Che affonda le sue radici nel pensiero di San Tommaso d'Aquino.

[24] Il riferimento è al noto esperimento mentale del carrello ferroviario, elaborato da P.R. FOOT, *The Problem of Abortion and the Doctrine of the Double Effect*, in *Virtues and Vices and Other Essays in Moral Philosophy*, cit., p. 24 ss., in part. p. 27. L'elemento di curiosità è rappresentato dalla circostanza per cui

proprio i portatori di interessi a comprendere come, quando si parla di etica dell'intelligenza artificiale, si intende fare riferimento alla capacità di "scegliere".

Nel binomio "algoritmo etico", d'altronde, gli scienziati sono consapevoli che il ruolo della regolamentazione sia cruciale, soprattutto perché spetta alle scienze umane e sociali il compito di definire il campo d'azione degli algoritmi, senza dimenticare, però, che gli approcci puramente normativi, basati su leggi e regolamenti, scontano un grosso problema: «*they don't scale*» [25]. Allora, il compito del legislatore, al netto dell'incommensurabilità dell'approccio regolamentare rispetto a quello "algoritmico", sembra essere quello di ricostruire la "genealogia" delle scelte che si manifestano nel solco della catena del valore dell'intelligenza artificiale, magari prendendo spunto dalla valutazione dei rischi reputazionali nell'ottica dell'attività d'impresa (anche a prescindere da determinati obblighi giuridici), vista la consolidata prassi dei codici "etici" d'impresa [26] e l'attenzione per la *Corporate social responsability* [27].

In fin dei conti, la sfida che la "nuova" scienza dell'"algoritmo etico" ha lanciato nei confronti del regolatore (non solo) europeo (che, per dirla con Nietzsche, a tratti è "umano, troppo umano") sembra essere opposta rispetto all'atteggiamento tecnologicamente neutro della normativa europea [28], al fine di non pregiudicare né la sperimentazione né, tantomeno, la capacità di incidere sulla stessa attraverso scelte "etiche".

[ATTILIO ALTIERI]

---

l'esperimento del *trolley problem* e la sua metodologia sono stati impiegati dal MIT di Boston per l'automobile a guida autonoma: v., più approfonditamente, https://www.moralmachine.net/.

[25] M. KEARNS, A. ROTH, *The Ethical Algorithm*, cit., p. 192.

[26] Tra i molti, S. ROSSI, *Luci e ombre dei codici etici d'impresa*, in *RDS*, 2008, 1, p. 23 ss.; più recentemente, A. SCOTTI, *I codici di condotta tra mercato, impresa e contratto*, Milano, 2019, p. 41 ss.

[27] Tra i molti, M.V. ZAMMITTI, *La responsabilità della capogruppo per la condotta socialmente irresponsabile delle società subordinate*, Milano, 2020, p. 48 ss. Oltretutto, il discorso sarebbe ancora più stringente nei confronti di quelle imprese che dichiaratamente avessero assunto impegni "etici" con i portatori di interesse.

[28] E v., da ultimo, la proposta di regolamento sui "prodotti macchina", COM(2021) 202 final, sempre del 21 aprile 2021.