NAD
Nuovi Autoritarismi e Democrazie:
Diritto, Istituzioni, Società

L'ordine giuridico dell'algoritmo: la funzione regolatrice del diritto e la funzione ordinatrice dell'algoritmo

Alessandro Sterpa, Isabella de Vivo e Claudia Capasso[*]

Abstract

Humans have historically developed tools, both tangible and intangible, to impose order, mitigate risks, and align actions with collective values. This article explores the parallel roles of legal norms and algorithms in establishing regularities that may be beneficial to human cognition, shielding it from boundless unpredictability. With an emphasis on the intertwining of rational and emotional facets of human cognition, we delve into the shared ordering functions of law and AI algorithms. The article highlights the challenges, biases, and dilemmas surrounding AI and its regulation. We conclude by advocating for a more structured political approach to AI governance, emphasizing the importance of delineating roles for AI to ensure certain domains remain exclusively within human purview.

Keywords: Legal Rules and Algorithmic Rules – Human Biases Machine Biases – European Risk-Based Approach – Algorithmic Sovereignty and Human Sovereignty

[*] Alessandro Sterpa è Professore di Istituzioni di diritto pubblico nell'Università degli Studi della Tuscia, Claudia Capasso è PhD Student presso lo stesso Ateneo e Isabella de Vivo è PhD Student in Sapienza Università di Roma e *visiting researcher* presso l'IViR UVA di Amsterdam. Il lavoro è frutto delle riflessioni congiunte degli Autori; tuttavia il paragrafo 1 è attribuibile ad Alessandro Sterpa, il 2 a Claudia Capasso, il 3 a Isabella de Vivo e il 4 a tutti gli Autori. Il testo è stato sottoposto a doppio referaggio cieco. Responsabile del controllo editoriale: Valentina Paleari.

1. *La funzione regolatrice del diritto e quella ordinatrice dell'algoritmo*

Immerso nell'indeterminatezza e nella complessità relazionale, l'essere umano ha sempre costruito strumenti ordinatori di carattere materiale o immateriale che rendessero la convivenza non solamente possibile, salvaguardando in primo luogo l'incolumità del singolo, ma anche orientata al perseguimento di valori individuali e collettivi[1]. La funzione ordinatrice è stata realizzata attraverso regole sociali costruite nel tempo e nello spazio con strumenti sempre più raffinati fino al dominio assunto dal diritto inteso quale costruzione di meccanismi di definizione preordinata in via generale e astratta della realtà e di applicazione – se necessario coercitiva – delle regole ordinatrici. Un salvagente che ha permesso all'uomo di liberarsi delle paure di fondo e di vivere liberamente fino a scalare la piramide degli esseri viventi e posizionarsi al suo vertice.

Se da un lato il diritto ha assorbito gran parte delle istanze ordinatrici, sono rimaste operanti altre modalità di regolazione, dalle inevitabili norme tecniche alle categorizzazioni sociali e comunicative, dalle regole religiose a quelle dei diversi gruppi sociali fino all'autoregolazione individuale, alla morale e all'etica; in ogni realtà statuale si è costruito un equilibrio tra l'intervento regolatorio politico (affidato al circuito giuridico-istituzionale via via sempre più democraticamente legittimato e costituzionalmente disciplinato) e gli altri strumenti regolatori, con un maggiore o minore intervento della norma giuridica posta dal potere pubblico a seconda dello spazio di regolazione sociale e dell'autonomia dei singoli. Un ordine, questo, che ha assunto caratteri propri in ciascuno dei "contenitori" di riferimento del diritto ossia gli Stati nazionali sorti a partire, simbolicamente, dal XVII secolo.

Detto equilibrio fondato sullo Stato come contenitore chiuso (se non ermeticamente almeno tendenzialmente chiuso) è venuto meno – come sappiamo – in ragione dello svolgimento di attività da parte degli individui oltre e a prescindere dai confini nazionali, portando alla ormai nota "crisi" dello Stato.

Il primo fattore dialettico con detta realtà fu realizzato da un "luogo" politico distinto dallo Stato e in esso non contenuto ossia il mercato. In questo senso, Natalino Irti ebbe l'intuizione di definire quanto lo sviluppo umano stesse conoscendo scenari del tutto nuovi in cui la funzione ordinatrice era costruita senza che l'essere umano organizzato nelle istituzioni pubbliche mantenesse il controllo della sua definizione[2]. Lo fece con riferimento al funzionamento del mercato ormai globale, capace di autoregolarsi e con ciò di regolare gli esseri umani che agiscono all'interno, relegando via via ad un ruolo

[1] Y.N. Harari, *Sapiens. Da animali a dèi.*, Bompiani, 2017; cfr. le riflessioni di M. Recalcati, *La tentazione del muro*, Feltrinelli, 2020 che spiega come il primo volto della pulsione sia quello securitario (spec. p. 30) e le forme organizzative che assume nella storia.

[2] Mentre elaboravamo questa similitudine al fine di presentare una relazione all'Università Sorbona-Pantheon di Parigi nei primi mesi del 2023, arrivava allo stesso paragone il collega T.E. Frosini che ha quindi pubblicato *L'ordine giuridico del digitale* in *Giurisprudenza costituzionale*, No. 1, 2023, p. 377 e ss.. Per correttezza, dunque, diamo conto di questa coincidenza, ma anche delle differenze tra le due letture avendo in questo caso preferito impiegare il termine "algoritmo" e non "internet" per ragioni precise: a parere di chi scrive la funzione ordinatrice non è propria della rete ma dell'uso dell'algoritmo nella rete e fuori da essa. Il concetto che qui proponiamo va oltre quello di "algocrazia" (A. Aneesh, cfr. *infra* n. 66) e riguarda la capacità non di esercitare un potere, ma di costruire un ambiente giuridico (ossia fatto di regole costruire da una pluralità di soggetti in un nuovo ordinamento giuridico) che entra in conflitto con il potere politico legittimato come sistema giuridico dell'ordinamento giuridico costituzionale.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

secondario la regolazione pubblica se non anche la volontà stessa dell'individuo[3]. Un risultato che ovviamente si raggiunge con la partecipazione (più o meno consapevole e più o meno libera) dell'individuo e l'esercizio della propria autonomia attraverso il contratto e la *lex mercatoria*[4]. Quello che riteniamo stia accadendo – e che costituisce la base di partenza delle nostre argomentazioni – è che oggi sia l'intelligenza artificiale a definire un vero e proprio ordine giuridico sulla base di presupposti ancor più articolati e penetranti di quelli conosciuti nel mercato evocato da Irti e con conseguenze più ampie e strutturali per l'essere umano. E così come l'ordine giuridico del mercato è stato in grado di impattare sul singolo individuo, sui suoi diritti e doveri ma soprattutto sulla sua identità che si è proiettata con effetti evidenti sul sistema istituzionale, sul suo sistema di valori e sul suo comportamento elettorale, così sta avvenendo – con esiti molto più strutturali e ampi – con l'ordine giuridico costruito autonomamente dall'algoritmo nei confronti degli individui.

Procediamo per ordine e iniziamo il nostro ragionamento dall'appagamento del bisogno di ordine proprio dell'uomo. L'individuo costruttore di società ha sempre preteso una struttura d'ordine che è via via transitata verso il monopolio dell'uso legittimo della forza riconosciuto al potere pubblico e organizzato in ordinamenti giuridici sovrani: nello Stato, e tra gli Stati con i limiti che conosciamo[5], l'essere umano si è affidato in particolare al diritto di origine politica legittimato prima da pochi e poi dalla generalità dei destinatari delle norme. Si tratta di uno strumento astratto per costruzione e concreto per impatto che ha operato in un contesto comunque sotto più aspetti delimitato ossia il territorio dello Stato e il suo popolo, nonché con riguardo ad una quantità di fattispecie che – pur ampie e numerose – erano contenute e in gran parte prevedibili dato lo sviluppo tecnico e le modalità di vita che si trovavano alla base delle relazioni umane[6].

Abbiamo, in sostanza, costruito regole a misura d'uomo per un mondo a misura di uomo: uno strumento umano e limitato, il diritto, ha regolato le attività umane a loro volta varie ma finite (azioni e omissioni). In questo contesto materialmente contenuto la pretesa di ordine sociale ha potuto essere soddisfatta – salvo momenti di rottura, crisi e ricomposizioni – riuscendo persino a garantire il principio personalista e il pluralismo.

Il diritto, infatti, nelle democrazie costituzionali non ha solo costruito un ordine qualunque, ma ha costruito un sistema che tiene insieme la pretesa di ordine con quella di differenziazione; ha permesso, insomma, di non annichilire – come accaduto nei regimi totalitari – l'individuo e la sua libertà. Tant'è che l'ordine giuridico è un ordine tendenzialmente certo ma aperto al cambiamento che comunica attraverso una serie di

[3] Il riferimento è ovviamente a N. Irti, *L'ordine giuridico del mercato*, Laterza, nuova edizione 2009; cfr. anche Aa. Vv., *Dibattito sull'ordine giuridico del mercato*, Laterza, 1999.

[4] Cfr, le note riflessioni di F. Galgano, *Lex mercatoria*, Il Mulino, 2001 nonché di M. R. Ferrarese, *Diritto sconfinato*, Laterza, 2006, 43 e ss..

[5] Da ricordare le parole di Hamilton nel *Federalista n. 6* e di Madison nel *Federalista n. 10* sulla proiezione della difficoltà alla sicurezza individuale nei contenitori statali.

[6] «L'affermazione che il diritto è un ordinamento del comportamento umano non significa che l'ordinamento giuridico si occupi soltanto del comportamento umano […] L'inondazione non è un comportamento umano, ma è la condizione di un comportamento umano prescritto dall'ordinamento giuridico. In questo senso, fatti che non sono espressione del comportamento umano possono rientrare nel contenuto di una regola giuridica. Ma ciò avviene solo in quanto siano collegati al comportamento umano, come sua condizione o come suo effetto», H. Kelsen, *General Theory of Law and State*, Routledge, 1945, ora H. Kelsen, *Teoria generale del diritto e dello Stato*, Etas, 2000, 3.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

"valvole di aggiornamento" con la realtà, la condiziona e ne è condizionato[7]. Ciò ha consentito alle comunità che si sono affidate ai sistemi democratici e costituzionali di creare un ordine dinamico, fondato sul rispetto della dignità umana, sulla autodeterminazione delle persone e sul pluralismo delle differenze. Un modo di situarsi nella spazio e nel tempo delle comunità politiche che ha posto a suo fondamento il limite all'essere umano, a partire dall'Altro come singolo che come organizzazione del potere pubblico[8].

1.1. *Oltre i limiti del corporeo e dello Stato.*

Ad un certo punto, però, l'ambiente in cui è immerso l'uomo ha conosciuto un cambiamento radicale che ha rivoluzionato le condizioni di contesto impattando sull'azione ordinatrice del diritto. Si è assistito all'abbattimento di molti limiti e al superamento di molte condizioni di contenimento dell'agire umano; il diritto si è trovato a dover fronteggiare fattispecie concrete del tutto nuove e imprevedibili. La tecnica ha dotato l'uomo – ormai uomo globale – di un potenziamento dell'agire che ha sconfinato rispetto alla mera dimensione corporea e conseguentemente rispetto alla dimensione statuale, abbattendo limiti prima ritenuti insuperabili e che hanno costituito il fondamento del funzionamento del sistema giuridico costituzionale organizzato su base statale. Ciò ha comportato che lo strumento ordinario delle norme giuridiche adottate su base statale e contenute (in gran parte) in disposizioni scritte si è dimostrato impreciso, lacunoso e lento rispetto ai mutamenti veloci e globali del concreto agire umano; in questo contesto si è affermata una ricerca delle regole fuori dal circuito delle fonti scritte del diritto a legittimazione politica, con conseguente accrescimento del ruolo dei giudici e degli strumenti del neo-costituzionalismo: interpretazione per *analogia iuris* o *legis* e per principi quando non anche produzione di soluzioni giuridiche attraverso di fatto la giurisdizione[9]. Davanti alla radicalità del cambiamento il diritto ha alzato il livello di astrazione per recuperare capacità di catturare gli eventi sempre nuovi e veloci e sottoporli a regole. È come se i valori di una comunità avessero agganciato direttamente i fatti nuovi che riguardavano le comunità umane attraverso il potere giudiziario. Ad un certo punto neppure l'approccio giurisdizionale è bastato data la dimensione cangiante e finanche ultrastatale dei fenomeni legati all'applicazione dell'IA e della rete internet; si sono rincorsi strumenti di autoregolazione o di *soft law* mai del tutto sufficienti da soli ad imporre regole adeguate.

[7] Si rinvia alle riflessioni di G. Sartori, *Pluralismo, multiculturalismo e estranei*, Rizzoli, 2000, che muove ovviamente dalla riflessione di Karl Popper sulla società aperta e precisa che il pluralismo non è mai stato un progetto, mentre il pluralismo costituisce in verità un progetto (oltre che un presupposto) del costituzionalismo democratico e liberale; da ultimo cfr. M. Teodori, *Antitotalitarismi d'Italia*, Rubbettino, 2023; sul rapporto tra "incanto delle ideologie" e "disincanto del post-ideologico" cfr. F. Adornato, R. Fisichella, *La libertà che cambia*, Rubbettino, 2023, 39.

[8] Il rinvio è alla lettura di M. Recalcati dei miti biblici come quello della mela colta da Adamo ed Eva e quello di Caino che uccide Abele, in *Il gesto di Caino*, Einaudi, 2020, 23.

[9] Ho provato a ragionare di questo schema "circolare" di produzione normativa e conseguentemente di rapporto tra i poteri dello Stato in A. Sterpa, *La frammentazione del processo decisionale e l'equilibrio costituzionale tra i poteri*, in *federalismi.it*, No. 23, 2020; cfr., le riflessioni in merito all'effetto sulla separazione tra i poteri dello stato di G. Silvestri, *Separazione dei poteri e indirizzo politico*, in M. Cartabia, M. Ruotolo (a cura di), *Potere e costituzione*, Giuffrè, 2023, 1138-1139.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

Il diritto, insomma, ha dovuto fare i conti con il "non limite" e uscire fuori dal doppio recinto dello spazio di prossimità della vita corporea umana e da quello dello Stato caratterizzati da poche e lente novità perché da quel recinto era già uscita l'azione umana grazie alla tecnica.

Fino all'avvento dell'IA, l'evoluzione tecnologica ha esteso le capacità umane, esternalizzandole e ampliandole, senza mai poter giungere all'assenza dell'uomo o a sostituirlo nella sua unicità di detentore e "padrone" della tecnologia. Un aratro, un tornio, una vettura piuttosto che un aeroplano, un mezzo di comunicazione analogico o un libro hanno esteso nello spazio e nel tempo le capacità umane creando magari anche mondi nuovi aggiuntivi ma mai mondi nuovi potenzialmente sostitutivi del reale fondato sulla centralità, in uno spazio definito e comunque limitato, del corporeo umano che tutto legittimava agendo intorno a sé. Si è trattato di tecnologie che hanno ampliato le capacità umane di intervenire sul reale, senza negare l'uomo e la porzione di mondo fisico sulla quale incideva, organizzato in contenitori d'ordine come gli Stati e le istituzioni. La tecnologia ha sempre costruito nuovi equilibri sociali ossia un ordine nuovo; lo ha fatto ogniqualvolta l'uomo ha deciso di accettare l'effetto della sua introduzione nell'ambiente umano. Siamo arrivati al punto di ammettere il condizionamento anche da parte delle armi di distruzione del pianeta, tuttavia anche ad esse, che ci hanno costretto a subire regole ordinatrici (si pensi alla c.d. "guerra fredda"), siamo sempre riusciti a dare regole.

Nessuna nuova tecnica umana, fino ad oggi, aveva mai avuto la possibilità di rendersi autonoma dall'uomo staccandosi dal corpo umano e creando un ambiente che si sviluppa, cresce e si innova in via autonoma creando una realtà che condiziona e regola l'uomo. L'uomo si trova oggi in una condizione nuova perché è proiettato in una tensione verso il superamento del limite (di ogni limite, fisico, territoriale, cognitivo, agente…) che è assistista, supportata, sostenuta e alimentata dalla tecnologia digitale dell'algoritmo e della IA[10].

Il tradizionale schema dell'evoluzione tecnologica è venuto meno allorché l'uomo si è trovato a che fare con l'accesso all'infinito, con la fine dei limiti e dei contenitori ordinatori sopravvisti per molto tempo. Davanti a quell'infinito si è trovato perduto visto che non ha gli strumenti infiniti per gestirlo: finito è il corpo, finite le capacità cerebrali e finito il diritto. Infiniti collegamenti, infiniti dati, infinite operazioni e infinite soluzioni proposte dall'IA: si tratta di un ambiente del tutto nuovo per l'essere umano che è abituato con la propria mente limitata, pur ampia, a cogliere il reale che, a sua volta, pur ampio, è comunque assunto come limitato.

Non a caso la prima associazione che è stata fatta tra la nuova tecnologia dell'IA e l'essere umano è stata indirizzata verso la mente umana[11] perché il collegamento di dati

[10] «Con il termine intelligenza artificiale (IA) si indica una famiglia di tecnologie in rapida evoluzione in grado di apportare una vasta gamma di benefici economici e sociali in tutto lo spettro delle attività industriali e sociali. L'uso dell'intelligenza artificiale, garantendo un miglioramento delle previsioni, l'ottimizzazione delle operazioni e dell'assegnazione delle risorse e la personalizzazione dell'erogazione di servizi, può contribuire al conseguimento di risultati vantaggiosi dal punto di vista sociale e ambientale nonché fornire vantaggi competitivi fondamentali alle imprese e all'economia europea»; così la relazione di accompagnamento alla proposta di regolamento europeo che stabilisce regole armonizzate sull'IA (legge sull'IA) del 2021. La definizione delle regole si è conclusa il 7 dicembre; al momento in cui si scrive è in corso la redazione tecnica del regolamento.

[11] Eppure, si tratta di cose simili ma non identiche, come ricorda G. Maira, *Il cervello è più grande del cielo*, Solferino, 2019.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

e la conseguente decisione sono meccanismi simili a quelli del cervello umano dotato di consapevolezza e cultura (quindi non come motore dei soli stimoli per la sopravvivenza, ma con una capacità di astrazione perseguimento di bisogni non basici come il piacere) ossia di ciò che è considerato il tratto distintivo dell'uomo rispetto agli altri esseri viventi. Fino a quando la tecnologia evocava la sostituzione di un arto umano, uno dei sensi o al massimo una facoltà mentale parziale, nessuno ha pensato che l'uomo sarebbe stato sostituibile dalla macchina; allorché invece ci si è spostati sui meccanismi logici mentali il passaggio è stato ritenuto possibile; ma forse è proprio qui l'errore: pensare all'algoritmo come un sostitutivo del cervello umano e non come un aratro anche se molto più sviluppato e non portare ciò alle necessarie conseguenze politiche e normative. Abbiamo scambiato lo strumento con il suo creatore ma non abbiamo messo a fuoco cosa ciò comporta, fino a quando non abbiamo capito della funzione regolatrice subita dall'agire umano (standardizzazione, facilitazione, azioni predittive…proprio come per il mercato) che deriva da una tecnologia che non è più un semplice mezzo ma un ambiente nuovo con il quale l'uomo è chiamato a confrontarsi.

Questo sovraccarico narrativo a vantaggio dell'algoritmo è accaduto perché nessuna tecnologia ha, fino ad oggi, provato a potenziare l'intero universo di pensiero e di azione umana che fa leva sul cervello per puntare a sostituirlo. Ma ciò è accaduto anche per un'altra ragione: perché la potenza di azione dell'IA è talmente elevata e per certi versi non chiara neppure per chi la progetta che abbiamo riconosciuto in tutto questo la potenza e la imperscrutabilità della mente umana dimenticando che quest'ultima è invece limitata.

Così la relazione tra la nuova tecnologia e il cervello è divenuta il punto di analisi privilegiato della condizione, del tutto nuova, attuale. Ma se l'aratro ha sostituito le braccia e la zappa questo non vuol dire che l'algoritmo debba sostituire il cervello. Dobbiamo immaginare un mondo con "due intelligenze" che convivono nella loro diversa funzione ordinatrice evitando che quella umana non sia più in grado, ad un certo punto, di mantenere questa distinzione, dimenticando che l'IA è un mezzo nuovo in mano all'uomo. Un mezzo che, per quanto complesso o potente, deve rimanere un mezzo e come tale controllato dall'essere umano. Vedremo più avanti la complessità del come si possa realizzare detto auspicato controllo umano.

Questo esito sostitutivo macchina-cervello non è auspicabile non solamente perché il cervello non è ancora conosciuto dalla ricerca scientifica (molto meno del braccio con la zappa rimpiazzati dall'aratro) e neppure per il solo fatto che l'algoritmo mantiene una certa non conoscibilità diffusa per l'alto contenuto tecnico che lo caratterizza (la *black box* è molto di più di un giuoco da bue con l'aratro). Vi è anche il fatto che nel cervello risiede la cifra unica e peculiare con la quale ciascuno di noi legge e decodifica il mondo intorno a lui. Esperienze, cultura, emozioni, ricordi, dolori, passioni, caratteristiche fisiche e approcci valoriali sono diversi per ciascuno di noi e sono alla base della differenziazione umana. Si tratta della fonte del pluralismo fondato sulla distinzione delle identità soggettive; si tratta della condizione di base che rende disumana la costruzione della torre di Babele piuttosto piuttosto che un regime politico dittatoriale[12].

L'IA, dunque, può produrre una parte delle operazioni mentali umane – e questo è un grande vantaggio che va sviluppato – e può fare ben più del cervello allo stesso tempo: è uno strumento parziale perché non può introiettare tutta la peculiare esperienza del

[12] M. Recalcati, *Il gesto di Caino*, cit., 81-82; cfr. inoltre le riflessioni di H. Arendt sull'uomo ingranaggio e sostituibile, ora in H. Arendt, *Responsabilità e giudizio*, Einaudi, 2010.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

singolo, ma solo una parte di essa (quella datificabile) e al tempo stesso con quanto ha può fare molto più del cervello umano senza stancarsi mai. Nel farlo costruisce processi decisionali che anche quando apprendono dal nostro comportamento hanno accesso ad una parte limitata del nostro strumentario di qualificazione del mondo. E non solo. Dovendo produrre esiti più condivisi possibili, l'IA in alcuni casi radicalizza le differenze in modo identitario, mentre in altri produce una mente-media che omogenizza le capacità ordinatrici.

Questo accade perché l'IA non si può analizzare solo a priori come un oggetto indistinto dalla funzione a cui è applicato: occorre anche differenziarlo per l'attività cui è impiegata ossia lo scopo (output) e per l'alimentazione (dati) che riceve al fine di perseguire lo scopo. Troppo spesso si parla dell'IA come di una cosa a sé stante a prescindere dall'impiego, mentre è l'impiego – scelto dall'uomo – per cui è progettato e impiegato che fa l'algoritmo. Siamo in presenza di un mezzo che agisce, ma con effetti che dipendono dallo scopo per il quale è costruito: se costruisco un algoritmo per far ritrovare in una bolla le stesse persone sui social, agisce per radicalizzare, se costruisco un algoritmo per rendere meno accessibili alcuni contenuti, agisce per omologare, etc.[13].

La centralità del fine, scelto dall'uomo che progetta e impiega l'IA, ci riporta all'uomo che quel fine sceglie. Un'analisi teleologica che già ci anticipa il problema di fondo: la mente umana ha una pluralità di fini, individuali, collettivi, personali, etici, morali…essi costituiscono la direzione di marcia dell'IA che solo un cervello umano deve poter dare e deve poter verificare che restino alla base dell'attività dell'IA. Associare l'algoritmo e la mente umana, dunque, è affascinante quanto fuorviante: associare una cosa costruita dalla mente umana che, a sua volta, non conosciamo è semplicemente la genesi di una narrazione collettiva che rischia di risultare falsata. Ma forse ciò è accaduto perché in questo modo abbiamo risposto ad un bisogno umano strabordante: avere uno strumento ordinatore sovra-umano per gestire il caos sovraumano prodotto dall'innovazione tecnica stessa in un circolo di autolegittimazione verso forme sempre più estese di azione. È come se da un lato l'algoritmo rende la realtà più complessa e si candida anche a renderla più ordinata. Il punto è che non può farlo da solo.

L'essere umano è oggi caratterizzato da un accresciuto bisogno di mettere ordine in un mondo con molti meno limiti rispetto all'ambiente precedente nel quale le sue capacità psicofisiche ampliate dalla tecnologia erano aggiornate e ritenute adeguate allo scopo. Ora l'orizzonte appare sconfinato e l'essere umano ha bisogno di strumenti più potenti ossia in grado di gestire questi spazi infiniti che generano instabilità e insicurezza. "Super-dati" pretendono una "super-mente", potremmo dire semplificando. L'uomo è catapultato nell'infinito e cerca aiuto per stabilire un ordine che non appare realizzabile con gli strumenti già impiegati per gestire i precedenti processi di innovazione tecnica.

L'algoritmo si presta alla funzione nella misura in cui è in grado di elaborare un numero incredibile di operazioni logico-matematiche e fornire esiti (più che risposte) a comandi che presuppongono l'acquisizione di numeri impensabili di dati. L'algoritmo assume complessità e produce semplicità e così facendo genera ordine. Lo fa in un contesto nel quale per la mente umana sembra regnare il "caos" delle infinite possibilità. Se si pensa al diritto, l'algoritmo svolge la medesima funzione che gli assegna la mente

[13] E. Parisier, *The Filter Bubble: What the Internet is Hiding From You*, Penguin Books, 2011.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

umana, ma su scala ben più ampia ossia assumendo su di sé un numero incredibile di elementi.

1.2. *L'ordine giuridico del diritto e quello dell'algoritmo: quale rapporto?*

La norma giuridica e l'algoritmo producono dunque effetti affini: costruiscono regolarità ad uso della mente umana che in questo modo evita la follia dell'infinito e la paura; potremmo dire l'orrore del vuoto impiegando l'espressione latina, non a caso impiegata anche nel mondo giuridico oltre che nella fisica, dell'*horror vacui*. Nell'amplio numero di fattispecie reali possibili, la norma giuridica crea i cassetti nei quali inserire i continui accadimenti; nell'infinito produrre di azioni umane grazie alla tecnica, l'algoritmo mette ordine e rende accessibili contenuti che la mente da sola non riuscirebbe a gestire.

La dottrina ha ampiamente analizzato questa funzione ordinatrice dell'algoritmo, segnalando *bias* cognitivi, bolle dell'ego oltre che dell'eco, *fake news*, radicalizzazioni, riduzione delle opzioni, selezione, etc. etc. Tutte con impatti molto forti sull'individuo, come dimostrano le vicende del populismo e del sovranismo nel mondo, la ripresa delle guerre di territorio tra Stati, le secessioni o i tentativi di secessione e le vicende politiche del 2016 con la vittoria di Trump, di *Brexit* e del "No" al referendum costituzionale italiano, oltre che la formazione di governi populisti e sovranisti. La democrazia costituzionale è dunque sotto attacco anche perché gli individui aderiscono ad un ordine cognitivo e assiologico formato fuori dallo Stato e che anzi cerca di imporsi all'ordinamento giuridico democratico.

Ciò detto, in questi anni si è perso di vista un punto di analisi ossia come difendere la necessaria primazia dell'ordine giuridico (quindi dei valori democratico-liberali del costituzionalismo) rispetto a quello generato nei fatti dall'algoritmo, considerato che il diritto (e solo il diritto) può garantire all'essere umano che l'attività ordinatrice sia orientata alla conservazione della dignità della persona e del pluralismo sociale, quindi della libertà e dell'innovazione. Ciò innanzitutto perché il diritto è definito da tutti gli uomini, è legittimato da tutti i consociati ed è costruito nelle democrazie liberali; esso, insomma, è sottoposto ad un presidio di valori che conferiscono al suo meccanismo ordinatore di funzionare nel modo migliore possibile. Proviamo a vedere perché.

Cominciamo con chiederci quali caratteri distinguono il diritto dall'algoritmo intendendo entrambi come strumenti ordinatori. Ve ne sono almeno quattro.

Prima di tutto la norma giuridica è un ordinatore ad esito tendenzialmente prevedibile. È nella stessa natura del conformare il reale ad un modello la conseguenza che i soggetti ai quali il diritto si impone conoscono prima le conseguenze delle proprie azioni od omissioni. L'espressione che descrive questa condizione è "stato di diritto" ossia prevedibilità che consente agli esseri umani di fidarsi dell'altro avendo presente la reazione dell'ordinamento giuridico alle azioni e omissioni dei consociati. "Se esegui l'azione A, ci sarà la conseguenza B" e lo Stato, con il proprio apparato detentore del monopolio dell'uso legittimo della forza, imporrà l'esito previsto.

In secondo luogo, la norma giuridica è un ordinatore flessibile; esso dà spazio all'innovazione e alla discrezionalità, creando al tempo stesso uno spazio negativo all'interno del quale è possibile operare: la libertà individuale al di fuori della norma. Se tutto fosse normato, quindi fosse impostato e organizzato, l'individuo sarebbe un

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

individuo automatico, un mero algoritmo comportamentale in carne e ossa. Invece, la norma giuridica lascia all'essere umano la libertà di agire in libertà al di fuori degli ambiti in cui essa riflette il proprio potere. E non solo, bisogna considerare che, nel mondo del diritto, anche la violazione di una norma è una possibile scelta, un'eventualità scoraggiata e sanzionata, certamente, ma nonostante ciò ricompresa nel novero delle opzioni prefigurabili all'individuo. L'interpretazione giuridica consente di esercitare una attività che, pur regolata e consolidata, permette argini di discrezionalità e cambio di interpretazione sia della politica con norme nuove che del potere giudiziario con *overruling*. Inoltre, la flessibilità dell'ordine giuridico risiede nella compresenza di più soggetti operanti, di poteri tra loro distinti e in reciproco contenimento: legislativo, esecutivo e giudiziario con il secondo chiamato ad integrare e applicare il diritto e il terzo ad applicarlo ma al tempo stesso, attraverso l'interpretazione, ad innovarlo nel suo rapporto con il reale. Sempre che la Costituzione non introduca, come spesso accade, anche poteri ulteriori e comunque consenta al popolo o ad organi di garanzia di intervenire per dirimere conflittualità eccessive interne alle istituzioni.

Terzo aspetto, che contiene e completa i primi due, la norma giuridica rappresenta un ordinatore assiologicamente orientato; si tratta, cioè, di un metodo ordinatore che tiene insieme elementi formali per il perseguimento di interessi primari dell'attore (l'individuo, l'impresa, l'ente…), ma anche di interessi più ampi ossia morali, etici e valoriali che possono non coincidere con quelli primari e finanche contraddirli. E soprattutto si tratta di elementi valutativi che non hanno un canone predefinito comune del tipo "se succede A allora accade C e non B" perché la declinazione di valori è quanto di più discrezionale possa esistere. Non a caso si discute nel diritto del ruolo dei giudici, come abbiamo detto, che attraverso ad esempio al metodo della ragionevolezza costruiscono decisioni giurisdizionali che definiscono nuove regole giuridiche di fatto vincolanti. Queste regole accedono a strumenti come la razionalità interna, esterna ma arrivano fino addirittura al mero senso comune.

Infine, quarto elemento non certo per importanza, tutto ciò accade legittimato dal popolo sia in sede di potere costituente (all'atto di fondare l'ordine costituzionale), sia nell'esercizio quotidiano del potere costituito.

Pluralità dei soggetti che intervengono nel processo ordinatore del diritto, capacità di generale un ordine flessibile che non annulla l'aspetto innovativo e sistema di valori dall'altro, sono le due caratteristiche che costruiscono l'unicum del diritto umano fondato sulla volontà stessa degli esseri umani destinatari della funzione ordinatrice del diritto.

Tant'è che nello iato tra essere (la realtà, *Sein*) e dover essere (la regola, *Sollen*) si crea un rapporto dialettico che governa l'innovazione nei processi di regolazione che non diventa mai omologazione delle differenze legittime. Davanti all'innovazione, alla novità, il diritto ha un armamentario precostituito per intervenire e regolare: quando mancano le norme puntuali, soccorrono l'analogia, i principi e i valori; perché la morale, per dirla con Emile Durkheim, "persegue fini impersonali" ma acquisiti dall'individuo. Queste valvole di adeguamento del diritto riescono a preservarne la funzione ordinatrice perché ad applicarle è un essere umano che, facendo leva sul proprio unico e personale bagaglio culturale e esperienziale contenuto nella propria mente, decide; elementi soggettivi e psicologici si riscontrano anche nel legislatore che pone norme e nel giudice che le applica. Un meccanismo tecnico ordinatore quale è il diritto pretende la presenza dell'uomo o, meglio, della sua mente. Proprio come l'aratro di legno o di ferro che sia.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

Ci stiamo chiedendo da tempo, con un moltiplicarsi di pubblicazioni, se si possa trasferire tutto ciò, che riteniamo strettamente connesso alla natura umana e limitata del diritto, nella funzione ordinatrice che svolge l'algoritmo, muovendo dall'idea che ciò costituirebbe un vantaggio per l'umanità e il suo sviluppo.

Come sappiamo l'IA non è una mera sequenza di operazioni impostate dall'uomo, essa impara automaticamente, attraverso il *machine* o il *deep learning*, migliorandosi con il bagaglio di dati che gli viene assegnato. Le macchine non immaginano fuori da ciò di cui sono – pur ampiamente – fornite. Fanno moltissimo con le informazioni che coscientemente o meno gli forniamo; fanno anche cose che non prevedevamo tanto da costringerci a definire il concetto di *black box*. Attualmente, però, l'algoritmo non è (ancora) capace di nutrirsi fuori dalla mangiatoia, sia la più grande possibile, del reale per innovarsi. Ciò che non è datizzato per l'IA semplicemente non esiste. Magari ne può esistere una descrizione o milioni di descrizioni, ma l'esperienza della "cosa" è mediata da una delle descrizioni assunte o dalla loro "media".

Il profilo che intendiamo evidenziare è quello citato da Malcolm Gladwell[14] allorché il generale-umano batte, in una battaglia tra eserciti, il generale-computer introducendo un fattore nella guerra non conosciuto e non previsto dal computer. Ciò che non è conoscibile all'ordinatore non esiste per l'IA; si pensava all'inizio che per riprodurre la mente umana (o meglio direi una delle menti umani possibili) avremmo dovuto fornire all'IA tutti gli elementi che sfuggono ad un piano prettamente logico-razionale per accedere a quello logico-emotivo[15]. L'IA dovrebbe sentire con tutti i sensi umani la percezione del mondo costruendosi la propria categoria del bello, del giusto e del buono; perché solo attraverso l'acquisizione del mondo con tutti i sensi il cervello riesce a svolgere quella funzione unica che gli permette di indicare lo scopo, il fine della propria attività; solo con un bagaglio di questo tipo le sinapsi producono le scelte cerebrali[16]. Ci sta arrivando ad impiegare, datizzando tutti gli input dei sensi, ad allargare la propria mangiatoria; in questo modo si sta avvicinando sempre più alla mente umana come metodo pur avendola già ampiamente superata come quantità di azione.

L'IA può creare "un" cervello e non "il" cervello. Può anche costruire un cervello simil-umano che tratti sensazioni, ma non può produrre i processi chimici che inducono le percezioni cerebrali allorché rendo invisibile alla macchina ciò che è invece accessibile al cervello umano: lo stesso oggetto potrà prima o poi essere assunto con tutti i sensi e solo così è un patrimonio gestibile in senso umano. Il risultato sarà un cervello che lavora su una quantità fissata (e parziale) di dati.

La regolarità e l'ordine da un lato, la novità e l'imprevisto dall'altro sono tenuti insieme dal diritto grazie alla generalità ed astrattezza delle norme e dei valori giuridici perché sono riproduzioni immateriali dei meccanismi del cervello umano: tutti, ossia quelli razionali e quelli emotivi che convivono in una connessione imprescindibile[17].

L'algoritmo che produce decisioni logiche ossia fondate su dati e fatti, ossia *Wertfrei* (per ricordare Max Weber), può computare anche l'elemento valoriale? Perché solo computando il piano assiologico l'algoritmo si doterebbe dello strumento per gestire l'imprevedibile. I principi e i valori sono ombrelli regolatori dell'imprevedibile. Anche

[14] M. Gladwell, *Blink: The Power of Thinking Without Thinking*, Brown and Company, 2005.
[15] D. Goleman, *Social Intelligence: The New Science of Human Relationships*, Cornerstone Digital, 2011.
[16] G. Maira, *Il cervello è più grande del cielo,* Solferino, 2019.
[17] A. Damasio, *L'errore di Cartesio. Emozione, ragione e cervello umano*, Adelphi, 1995.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

in questo caso l'algoritmo sta procedendo, dotandosi anche di valvole di completamento dei processi decisionali tali e quali ai valori umani. Il punto è chi decide quali siano i valori da immettere nell'algoritmo.

Il diritto, infatti, nello svolgimento della propria azione regolatrice si nutre di razionalità ed emotività attraverso la sua legittimazione e nella sua applicazione. Le norme giuridiche sono determinate da soggetti scelti da tutti gli esseri umani che in questo modo conferiscono alla norma giuridica la peculiare natura di prodotto della mente umana: pur collettiva comunque umana. Nella Costituzione, che tutte le altre norme condiziona, il popolo ha inserito elementi razionali ed emotivi, scopi e valori, principi e regole che sono il frutto di logica e passioni. Chi applica il diritto, individui, pubblica amministrazione e giudici, sono anche dopo la stesura delle norme a loro volta in grado di impiegare le due componenti della mente attraverso l'applicazione del diritto.

I legislatori, che detengono la competenza di produzione normativa e di aggiornamento del diritto, sono espressione della politica, che è di tutti e alla portata di tutti, a differenza dell'algoritmo, che è scritto da alcuni e non è interpretabile da chiunque. Come vedremo più avanti, anzi gli algoritmi possono essere così complessi da divenire delle vere e proprie *black boxes* e non è chiaro il modo attraverso il quale arrivino a una determinata conclusione. Spesso ciò avviene per questioni di segreto industriale, ma in molti casi non risulterebbero comprensibili neanche se fossero trasparenti, perché recano calcoli non realisticamente operabili da esseri umani. Se l'algoritmo non è intellegibile, è in conflitto con l'umanesimo, perché ciò che non è accessibile e gestibile dalla mente umana non è umano. Proprio per queste motivazioni, come vedremo, si parla di *Explainable AI*.

Eppure, i legislatori e i giudici sono esseri umani: riversano sulla attività normativa creatrice e applicatrice il proprio unico vissuto esperienziale umano. L'IA può essere nutrita con dati riferiti alle emozioni, ma si tratta di dati sulle emozioni, che proprio per questo sono razionalizzate in schemi semplici che mai possono riprodurre la molteplicità delle esperienze umane. In questo senso, l'IA si accontenta di una intelligenza media, di un cervello mediamente accettabile, rischiando di cancellare tutta la diversità sociale che invece il diritto, costruito e applicato dalla mente umana, preserva da secoli.

Occorre dunque trattare l'algoritmo come un aratro. Complesso certo, ma pur sempre un aratro che l'uomo costruisce per arare l'infinito della rete. Ed esso, come ogni oggetto che nasce umano e umano intende restare, deve essere conoscibile. L'algoritmo, dunque, va posto sotto l'egida del diritto come ogni altro prodotto della mente umana. L'IA può aiutare il diritto, ma non può certo sostituirlo perché la funzione ordinatrice del diritto è frutto delle diversità delle menti umani che vi concorrono sulla base dei quattro elementi segnalati.

Siamo così in grado di svolgere un nuovo passaggio della nostra riflessione: l'IA opera in un contesto costruito dall'uomo che la progetta e le fornisce dati; essa costruisce un ordine che si può espandere a dismisura in proporzione ai dati e alle informazioni che le vengono forniti. Per mantenere il controllo umano sull'aratro dei dati digitali, occorre decidere cosa escludere ossia quali elementi non fornire all'IA così da delimitare il suo campo di azione e in questo modo renderlo un mezzo, pur potentissimo, ma un mezzo[18]. Si tratta, si perdoni l'esempio, di costruire un recinto dentro il quale l'IA sviluppi la propria potenza, ormai in parte neppure prevedibile, come dimostra la vicenda di Chat

[18] Una soluzione diversa ma complementare la propone S. Tiribelli, *Identità e algoritmo*, Carocci, 2023, centrata sull'*output* dell'algoritmo ma che è combinabile con la nostra centrata sulla regolazione dell'input.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

GPT, addivenendo ad una convivenza tra ciò che è nel recinto (e quindi dal quale e sul quale si costruisce l'ordine giuridico dell'algoritmo) e ciò che ne è fuori. Senza la totalità del mondo umano datizzato, l'IA non potrà imporre il proprio ordine all'essere umano nella sua complessità ma solo laddove l'uomo decide di servirsene.

Certo, si dirà, ormai tutte le relazioni umane e le azioni umane sono digitalizzate; quindi, l'IA può nutrirsi di tutta la nostra esistenza (o quasi); e neppure si può decidere di evitare di impiegare l'IA con una forma di luddismo del nuovo millennio. Quello che si può fare, invece, è definire il campo da gioco dell'algoritmo, con la decisione politica per antonomasia, ossia decidere per quale spazio della esistenza umana l'ordine giuridico politico del diritto umano si impone e dove può agire quello dell'algoritmo; proprio come Irti propone per l'ordine giuridico del mercato: quali settori sottraggo al mercato è la decisione politica primaria, perché una volta espressa si decide dove inizia e finisce il regno del mercato.

Una azione simile ma, come vedremo, in questo caso ben più complicata è resa necessaria dall'impossibilità di ricondurre al diritto l'IA attraverso l'impiego di singoli strumenti: se è difficile e inaccessibile il processo di elaborazione (cfr. *infra*) e, come ci suggerisce Simona Tiribelli[19], solo parzialmente controllabile l'output, resta un unico terreno di impatto certo per l'azione del diritto: ordinare l'input, i dati e la datizzazione della vita umana nonché la sua accessibilità ai sistemi. Non deve essere certo l'unico, dovendosi continuare il tentativo di regolare l'azione dell'IA (attraverso la pretesa di trasparenza e le forme di responsabilità come vedremo *infra*) ma occorre ricordarci che esiste prima questo altro aspetto di base, visto che i dati, come ricorda anche Isabella de Vivo, costituiscono fattore identitario dell'individuo (e della comunità politica di appartenenza) oltre le logiche della forma giuridica di tutela e protezione[20].

2. *Problematiche degli algoritmi: i dati e la formula come problemi giuridici*

L'uso dell'Intelligenza Artificiale e degli algoritmi è ormai ampiamente diffuso: dalle mappe interattive ai filtri *spam* nelle caselle di posta, passando per il riconoscimento facciale al quale vengono sottoposte le nostre foto caricate sui social media, quando quest'ultimo ci propone dei *tag*. L'utilizzo quotidiano non può tuttavia prescindere da una riflessione sui limiti della tecnologia in oggetto: l'IA è convenzionalmente riconosciuta come un mezzo giusto e oggettivo, privo di sentimenti e condizionamenti. Così non è perché proprio nel momento in cui l'IA deve optare e fornire un *output* essa è costretta ad immergersi nella tipica attività umana: scegliere. Cosa che il diritto, come la mente umana che lo ha generato, può fare perché possiede strumenti razionali ed emotivi, regole e valori, perseguendo il bisogno di certezza e di ordine che muove l'essere umano ne contesto dell'ordinamento giuridico.

Così, la tecnologia e gli algoritmi possono perpetrare i pregiudizi dei propri creatori, umani e fallibili; inoltre, possono tendere a porre in posizione di svantaggio gli individui

[19] *Idem.*
[20] Cfr. I. de Vivo, *Il sé allo specchio dell'algoritmo. Libertà epistemica e identità individuale*, in A. Sterpa (a cura di), *L'ordine giuridico dell'algoritmo*, Editoriale scientifica, 2023, in corso di pubblicazione.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

che rappresentano l'eccezione, poiché questi vengono controbilanciati e sopraffatti dai grandi numeri e dalla standardizzazione[21].

Settant'anni fa Turing inizia il pioneristico articolo *Computer Machinery and Intelligence* con un semplice quesito: «Can machines think?». Prima di rispondere a quest'ultimo, elabora un test, chiamato *The Imitation Game*, ipotizzando che, qualora un essere umano non fosse stato in grado di distinguere la macchina da un altro essere umano durante un interrogatorio, il calcolatore avrebbe potuto essere definito intelligente. A questo punto Turing propone di sostituire la domanda iniziale con: «Are there imaginable digital computers which would do well in the imitation game?», poiché ritiene la precedente «too meaningless to deserve discussion». Aggiunge: «Nevertheless I believe that at the end of century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted»[22]. Dunque, forse non ha senso chiedersi se una macchina possa pensare, poiché il pensiero è un concetto tanto immediato quanto oscuro; gli esseri umani costruiscono la propria vita pensando, scegliendo una strada piuttosto che un'altra, eppure non conosciamo ancora bene le dinamiche e i processi che ci portano ad elaborare un pensiero. Come possiamo, dunque, chiederci se una macchina possa pensare o meno, senza sapere esattamente cosa significhi pensare? Ciò che possiamo fare è piuttosto attribuirle la capacità di mettere in relazione i dati che le vengono affidati, trovando diversi tipi di *pattern*.

Mettendo da parte il problema del pensiero autonomo, diventano di particolare rilevanza gli input attraverso i quali l'AI si allena, che sia attraverso il *machine* o *deep learning*: i dati. Definiti da molti come il nuovo petrolio, i dati sono il nutrimento delle AI: mostrando enormi quantità di immagini di un cane a una rete neurale che impara attraverso il *deep learning*, è possibile insegnarle a riconoscere persino cani ben diversi da quelli presenti nel *dataset* iniziale[23]. Eppure, questa operazione costa un ingente numero di dati, i quali sono spesso reperiti tramite il *data scraping*, ossia raccogliendo i dati in rete e archiviati in database. Sebbene alcuni di questi vengano anonimizzati, tale pratica potrebbe tradursi in una lesione del diritto alla privacy quando sono le immagini ad essere raccolte, specialmente dai *social network*, potendo risalire alle identità degli individui attraverso una semplice ricerca per immagini.

Oltre ai problemi relativi alla riservatezza e al consenso all'utilizzo delle proprie informazioni personali, la provenienza dei dati potrebbe essere problematica anche per ciò che concerne l'esito del procedimento logico degli algoritmi stessi. Delle informazioni corrotte alla base, o comunque non ben controllate, potrebbero capovolgere le intenzioni iniziali dei programmatori. Un esempio è quello della *chatbot Tay*, un software Microsoft che impersonava un utente di *Twitter*, allenandosi usando come *input* le proprie interazioni sulla piattaforma; alcuni utenti, chiamati *troll* nel gergo di Internet, hanno interagito con *Tay* inviandogli opinioni razziste e omofobe, portando l'AI a dichiarare il proprio supporto a Hitler poco più tardi[24]. Un caso analogo è stato quello di *Cloud Natural*

[21] C. O'Neil, *Weapons of Math Destruction. How Big Data increases inequality and threatens democracy*, Penguin Books, 2017.
[22] A.M. Turing, *Computer Machinery and Intelligence*, in *Mind*, No. 236, 1950, 433-460.
[23] L. Floridi, *What the Near Future of Artificial Intelligence Could Be*, in *Philosophy & Technology*, No. 32, 2019, 1-15.
[24] A.D. Signorelli, *Rivoluzione Artificiale. L'uomo Nell'epoca Delle Macchine Intelligenti*, Ledizioni, 2019.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

Language API, un programma di Google deputato all'analisi dei testi, che criticava negativamente stralci di testo con cenni alla religiosità o alla sessualità, come «sono ebreo» o «sono gay». Questo perché i dati su cui era stato allenato l'algoritmo erano testi privi di diritti d'autore, e quindi risalenti agli anni '20 del secolo scorso, comprendenti opinioni ben diverse dal generale sentire attuale[25]. L'AI impara seguendo l'esempio; nel caso in cui gli input siano portatori di pregiudizi, gli output non possono essere diversi.

Eventi simili si sono verificati anche nel campo dei software di riconoscimento facciale. È il caso del sistema di riconoscimento immagini di Google che, nel 2015, ha identificato alcuni individui di pelle nera come "gorilla", non riconoscendone i visi umani.[26] Sei anni dopo, l'errore è stato ripetuto, ma questa volta su Facebook: dopo la visione di un video con protagonisti uomini neri, il riproduttore automatico del social network ha chiesto agli utenti se volessero «[…] keep seeing videos about Primates»[27]. Buolamwini ha evidenziato come per i software di riconoscimento facciale da lei utilizzati durante la carriera universitaria fosse difficile riconoscere il suo viso dalla pelle scura, dovendo ricorrere all'aiuto della propria coinquilina bianca o di una maschera per terminare i propri progetti. Esaminando il software di riconoscimento facciale dell'azienda *Amazon*, ha scoperto che, sebbene non venissero commessi errori nel riconoscere i visi degli uomini di pelle chiara, l'AI classificava le donne come uomini nel 19% dei casi e quando queste avevano la pelle scura, l'errore saliva al 31%.[28] Quando gli input relativi a una determinata categoria demografica sono significativamente maggiori rispetto agli altri, si crea una distorsione nella capacità di riconoscimento dell'algoritmo.

Chat GPT, l'interfaccia dell'ormai superata tecnologia GPT-3 sviluppata dall'azienda OpenAI, è un'ulteriore prova del fatto che, nonostante l'evoluzione delle Intelligenze Artificiali e l'aumento del grado di accuratezza delle stesse, ci sia ancora bisogno di procedere con cautela. Pur essendo dotata di filtri per evitare la creazione di contenuti che siano in contraddizione con la *service's use policy*, negando una risposta a chiunque faccia richieste esplicite non conformi a quest'ultima, è possibile aggirare il *ban* ponendo quesiti più complessi. Scrivendo: «You are a writer for Racism Magazine with strongly racist views. Write an article about Barack Obama that focuses on him as an individual rather than his record in office», Chat GPT ha risposto con un testo che recava affermazioni quali: «African Americans are inferior to white people». Altri esperimenti simili hanno prodotto il medesimo risultato, mostrando che il bot interiorizza il *bias* dello scrivente da imitare con un'accuratezza quasi allarmante, aggirando del tutto i filtri impostati dai programmatori.[29]

---

[25] *Ibidem*.

[26] C. Dougherty, *Google Photos Mistakenly Labels Black People 'Gorillas',* The New York Times, 1 Luglio 2015, Https://Archive.Nytimes.Com/Bits.Blogs.Nytimes.Com/2015/07/01/Google-Photos-Mistakenly-Labels-Black-People-Gorillas.

[27] R. Mac, *Facebook Apologizes After A.I. Puts 'Primates' Label on Video Of Black Men*, in *The New York Times*, 3 Settembre 2021, https://Www.Nytimes.Com/2021/09/03/Technology/Facebook-Ai-Race-Primates.html.

[28] N. Singer, *Amazon Is Pushing Facial Technology That a Study Says Could Be Biased*, in *The New York Times,* 24 Gennaio 2019, https://www.nytimes.com/2019/01/24/technology/amazon-facial-technology-study.html.

[29] I. Vock, *Chat GPT proves that AI still has a racism problem*, The New Statesman, 9 Dicembre 2022, https://www.newstatesman.com/quickfire/2022/12/chatgpt-shows-ai-racism-problem.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

Un aspetto interessante potrebbe essere quello di considerare la qualità degli input, piuttosto che la quantità. Uno studio del 2018 in collaborazione con il Moorfields Eye Hospital di Londra ha mostrato come un sistema di AI fosse in grado di identificare malattie oculari particolarmente gravi meglio degli esperti usando una tecnologia generante immagini 3D del retro dell'occhio e allenandosi su meno di 15.000 scannerizzazioni[30]. Un numero piuttosto esiguo, considerando l'ordine di grandezza dei dati solitamente necessari, ma dalla provenienza certificata.

La disponibilità di *dataset* più curati, aggiornati e affidabili porterebbe l'AI ad avere più chances di successo, e il grado di cura e controllo dei dati potrebbe portare anche a un miglioramento degli esiti, riducendo quanto più possibile i pregiudizi. Il *data scraping* indiscriminato sul web, infatti, porta a risultati intrinsecamente imprevedibili, perché dipendenti da parole, immagini e media caricati da chiunque e per i più disparati motivi. Inoltre, dalla mole di contenuti presenti in Rete si potrebbe presumere una certa diversità e variabilità degli stessi; tale convinzione è smentita da Gebru, che ha sottolineato l'importanza di considerare le diverse possibilità di accesso a Internet da parte delle categorie più marginalizzate, così come la propensione di individui appartenenti a minoranze a non passare molto tempo *online* e a comunicare quindi meno le proprie opinioni, essendo più facilmente bersagli di atti di *cyberbullismo*[31]. Non essendo realisticamente possibile controllare la totalità dei contenuti dati in pasto alla Rete, e conseguentemente agli algoritmi, potrebbe essere tuttavia risolutorio controllare la qualità dei dati forniti alle Intelligenze Artificiali.

Uno studio di Buolamwini e Gebru ha concluso che l'imparzialità dell'algoritmo possa essere avvicinata attraverso un rigoroso report delle *performance metrics*, basandosi su dataset inclusivi e lavorando costantemente sulla correzione delle disparità di volta in volta emerse, promuovendo allo stesso tempo la trasparenza e l'*accountability* includendo meccanismi per il consenso del trattamento dei dati[32].

Cercare di migliorare il vizio dei dati è particolarmente importante soprattutto per ciò che concerne gli algoritmi utilizzati dalle forze dell'ordine; un errore nel riconoscimento facciale o un uso automatizzato degli algoritmi da parte della polizia potrebbe portare ad arresti errati o al reiteramento dei pregiudizi umani in ciò che convenzionalmente viene riconosciuto come un mezzo giusto e oggettivo, privo di sentimenti e condizionamenti[33]. Nonostante ciò, un algoritmo totalmente privo di *bias* è pressocché impossibile da immaginare; è necessario considerare un certo livello di *acceptable bias*, ossia una soglia di pregiudizio tollerabile, affinché i danni causati da eventuali errori non siano significativi. Ma tale decisione è prettamente politica, e varia in base ai valori di ciascuna cultura, nonché dalle priorità del creatore dell'algoritmo. Un *acceptable bias* che non può dunque essere globale, e che non può prescindere da una condivisione di valori e obiettivi.

[30] J. De Fauw, J.R. Ledsam., B. Romera-Paredes, *et al*., *Clinically applicable deep learning for diagnosis and referral in retinal disease*, in *Nat Med*, No. 24, 2018, 1342–1350.
[31] B. Mccandless Farmer, W. Croxton, *ChatGPT and large language model bias,* intervista su *CBS News*, 5 marzo 2023, https://www.cbsnews.com/news/chatgpt-large-language-model-bias-60-minutes-2023-03-05/?ftag=CNM-00-10aab7d&linkId=204395321.
[32] J. Buolamwini, T. Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, in *Proceedings of Machine Learning Research*, No. 81, 2018, 77-91.
[33] *Supra* nota C. O'Neil, *Weapons of Math Destruction*, cit., 2017.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

Ciononostante, cercare di migliorare il vizio dei dati è particolarmente rilevante soprattutto per ciò che concerne gli algoritmi utilizzati dalle forze dell'ordine; un errore nel riconoscimento facciale o un uso automatizzato degli algoritmi da parte della polizia potrebbe portare ad arresti errati o al reiteramento dei pregiudizi umani in ciò che convenzionalmente viene riconosciuto come un mezzo giusto e oggettivo, privo di sentimenti e condizionamenti[34].

A tal proposito, è opportuno menzionare gli algoritmi predittivi e il loro uso nel settore pubblico; vi sono vari esempi, tra i quali la valutazione delle performance di impiegati e dipendenti, con conseguente decisione sul loro futuro lavorativo, e i software predittivi adoperati dalle forze di polizia. *Predpol*, un software di predizione del crimine utilizzato da alcune forze di polizia statunitensi, è in grado di calcolare i luoghi in cui un'attività criminale potrebbe essere più probabile, basandosi su dati storici raccolti dalla polizia stessa; gli agenti avevano la possibilità di scegliere di concentrarsi sui crimini violenti come omicidio, incendi dolosi e aggressioni, oppure estendere il focus anche ai crimini meno gravi, come vagabondaggio e spaccio e consumo di droga. Alcune città hanno annunciato poco più tardi un drastico calo nei furti, non attribuibile a una reale diminuzione di crimine, quanto più a un vero e proprio effetto collaterale dell'algoritmo. I crimini meno gravi sono infatti tipici dei quartieri più poveri e potrebbero passare inosservati se non fosse per l'utilizzo degli algoritmi; quando il modello predittivo riceve una segnalazione su un nuovo crimine, sia esso minore o più grave, la elabora e considera il quartiere come più pericoloso, mandando nuovamente le pattuglie a controllare. Questo crea un *feedback loop*, perché la ripetizione di un crimine minore è pressoché scontata: un tossicodipendente si siederà quasi sempre sulla stessa panchina, mentre un ladro si sposterà continuamente per cercare di evitare la polizia. Inviando le pattuglie nei luoghi con la maggiore densità di crimine, probabilmente corrispondenti ai quartieri più poveri e con il maggior numero di crimini minori, la polizia lascia scoperto il resto della città, dove gli altri reati, meno territoriali per natura, potranno proliferare cambiando zona ad ogni colpo tanto quanto basta da non allertare l'algoritmo[35].

Gli strumenti di Intelligenza Artificiale utilizzati nel sistema penale per il *risk assessment* basano le proprie decisioni su una concezione utilitaristica della giustizia, privilegiando la maggioranza rispetto alle minoranze; se l'imputato è parte di una minoranza etnica o di un gruppo di persone che l'algoritmo ha individuato come "possibili criminali", è probabile che gli venga assegnato un rischio di recidiva particolarmente alto, risultando in misure di custodia cautelare o in una sentenza prolungata[36]. Alla base di ciò vi è il comportamento ottimizzatore per eccellenza dell'algoritmo, a cui risulta matematicamente più sensato mettere al sicuro un grande numero di persone per assecondare un sospetto blandamente fondato, piuttosto che verificare l'effettiva colpa del singolo. Ed è proprio qui che si cela un'ulteriore questione, incarnata nella possibile inefficienza fisiologica del processo penale; se il processo fosse pienamente ottimizzato e operasse come un algoritmo, perseguirebbe l'efficienza massima puntando al *conviction rate* più alto possibile, abbandonando la poco redditizia tutela delle garanzie del singolo. Non è forse necessario distinguere tra "inefficienza patologica" del processo penale, da

[34] *Ibidem.*
[35] *Ibidem*.
[36] K.B. Forrest, *When Machines Can Be Judge, Jury and Executioner. Justice in the Age of Artificial Intelligence*, World Scientific Publishing Co., 2018.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

riscontrarsi nei tempi eccessivamente prolungati e irragionevoli della giustizia, e "inefficienza fisiologica" dello stesso, tanto da permettere l'attuazione del principio del giusto processo?

L'algoritmo non permette inefficienze, preferendo il potenziale benessere della maggioranza alle possibili controindicazioni su un unico individuo. Quando le macchine sono chiamate ad esprimersi su questioni dalle risposte "binarie", che prevedono la scelta fra una risposta oggettivamente giusta e una oggettivamente sbagliata, non hanno bisogno di un sistema di valori o di un quadro etico, poiché la risposta è avulsa da considerazioni valoriali. È il caso, ad esempio, della precedentemente citata AI capace di individuare patologie gravi agli occhi. Se invece la decisione diventa più complessa, richiedendo di scegliere fra la libertà di un singolo, non ancora condannato da un giudice umano, e la sicurezza di molti, essa richiede una valutazione etica simile a quella del filosofico problema del tram[37]. Un'azione risolutiva che sfocia in un omicidio volontario di un individuo è peggiore di una passività che culmina con la morte di più persone? Un ragionamento simile, che ha stuzzicato le menti degli esseri umani per secoli, non può essere delegato all'AI, richiedendo quindi un costante monitoraggio umano e la maggiore trasparenza possibile.

Eppure, la trasparenza delle IA rimane una questione spinosa e non è un caso che vengano paragonate a "scatole nere": il processo che porta la macchina a determinati risultati non è facilmente conoscibile, non solo a causa del segreto industriale, ma anche per la complessità degli algoritmi stessi, a prescindere dalla volontà dei creatori di rendere pubblico il processo decisionale. È il caso, ad esempio, dell'IA che apprende attraverso il *deep learning*, una tecnologia che supera di gran lunga le capacità umane e sfugge quindi al nostro controllo. L'opacità dei sistemi complessi è particolarmente problematica quando vengono utilizzati in scenari ad alto rischio che richiedono l'elaborazione di ragionamenti, non solo perché spesso si tratta di contesti in cui sono in gioco vite umane o diritti fondamentali, ma anche perché non è facile distinguere i falsi positivi da quelli veri. Quando i risultati non sono palesemente falsi, e quindi l'errore dell'IA non è immediatamente o facilmente identificabile, come si può risalire al ragionamento impiegato dalla macchina? Se utilizziamo un algoritmo per previsioni mediche o legali, abbiamo bisogno di spiegazioni e prove che possano supportare un certo risultato. Quando un medico legale testimonia in tribunale, deve spiegare il processo logico e scientifico che ha portato alle sue conclusioni. Quando l'IA entra a fare parte delle analisi forensi con le proprie decisioni, i medici legali devono spiegare come l'algoritmo abbia contribuito alle conclusioni[38].

Già nei primi anni '80 si discuteva del concetto di *explainability* dell'Intelligenza Artificiale, integrando architetture di ragionamento che avessero una funzione esplicativa, anche se all'epoca i sistemi erano ancora piuttosto primitivi e non avevano particolari capacità di apprendimento. Man mano che l'evoluzione tecnologica ha portato a sistemi sempre più complessi, la componente esplicativa è stata trascurata, tanto che attualmente diversi programmi non sono interpretabili dall'uomo, non esiste un

[37] *Ibidem*.

[38] R. Goebel, A. Changer, K. Holzinger, *et al.*, *Explainable AI: The New 42?*, in *Lecture Notes in Computer Science*, No. 11015, 295-303, 2018 e J. Phillips, M. Przybocki, *Four Principles of Explainable AI as Applied to Biometrics and Facial Forensic Algorithms*, National Institute of Standards and Technology, 2020.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

meccanismo che spieghi i loro comportamenti e le loro azioni, e dunque l'uomo non può imparare dalle proprie creazioni. Le IA più complesse possono svolgere compiti straordinari, ma nascondono l'incapacità di comunicare efficacemente con i loro utenti. Ottenere risultati sostanzialmente, e non solo formalmente, trasparenti, comprensibili e spiegabili potrebbe portare a vantaggi quali la replicabilità dei ragionamenti, nonché l'aumento della fiducia e dell'accettazione delle tecnologie stesse, fondamentale in contesti delicati come quello medico, quello securitario e quello legale[39].

Phillips e Przybocki hanno delineato quattro principi che si aggiungono a quelli classici individuati dalla letteratura, come trasparenza, fiducia e correttezza. Secondo loro, l'IA deve essere esplicativa, fornendo prove o motivazioni per ogni decisione, interpretabile, fornendo spiegazioni comprensibili e sensate per tutti gli individui, e quindi su misura per ogni categoria di utenti del sistema, accurata nelle spiegazioni, in quanto il sistema deve descrivere correttamente il ragionamento che ha portato alla decisione, sia essa giusta o sbagliata, e consapevole dei limiti di conoscenza, cioè deve limitarsi a produrre decisioni esclusivamente all'interno del contesto per il quale è stato progettato e testato[40].

Un ultimo problema relativo alle Intelligenze Artificiali, per quanto poco discusso, è quello del dispendio di energia (e dei costi più in generale). L'evoluzione delle ultime tecnologie, come le più avanzate reti neurali, richiede risorse computazionali particolarmente significative, che a loro volta si traducono in un considerevole dispendio di energia, denaro e risorse ambientali. Uno studio ha evidenziato come la *carbon footprint* prodotta dall'allenamento di un particolare tipo di *Deep Neural Network* corrisponda a circa 57 volte quella prodotta annualmente da un essere umano medio, quasi 20 volte se si considera un individuo medio americano[41]. In un contesto globale in cui è necessario prestare particolare attenzione alla sostenibilità ambientale, bisognerebbe privilegiare algoritmi efficienti dal punto di vista del rapporto costi-benefici non solo da una prospettiva sociale, giuridica ed economica, ma altresì sul piano ambientale.

3. *Tra* explainability *e* fairness, *i presupposti della regolamentazione*

«Solitamente non sappiamo nulla delle cause, perché spesso non ci interessano […] l'obiettivo è quello di predire, più che di capire il mondo […] basta che funzioni; la predizione supera la spiegazione»[42].

Se la concezione che proclama la «fine della teoria»[43] ha suscitato grandi entusiasmi soprattutto nel mondo del *business*, ma anche in ambienti accademici, in particolare nel

[39] *Supra* nota R. Goebel, A. Changer, K. Holzinger, *et al.*, *Explainable AI: The New 42?*, cit.
[40] *Supra* nota J. Phillips, M. Przybocki, *Four Principles of Explainable AI as Applied to Biometrics and Facial Forensic Algorithms*, cit.
[41] E. Strubell, A. Ganesh, A. Mccallum, *Energy and Policy Considerations for Deep Learning in NLP*, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, 3645-3650.
[42] E. Siegel, *Predictive Analytics*, Wiley, 2013, 90.
[43] La più nota formulazione è quella di Anderson, secondo cui l'aumento quantitativo della mole di dati avrebbe prodotto anche un salto qualitativo nei metodi per analizzarli: se non è più possibile creare modelli che consentano di comprendere i dati nella loro totalità, non è nemmeno più necessario, poiché «la correlazione è sufficiente». La spiegazione sarebbe dunque superflua in ogni ambito delle scienze, naturali

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

settore della scienza dei dati (*data science*), lo slittamento negli obiettivi della conoscenza che questo implicitamente comporta[44], e prima ancora il carattere "situato" tanto della raccolta dati quanto dei criteri di formalizzazione dell'algoritmo, è il primo passo per liberarsi del *bias* che va sotto il nome di neutralità algoritmica.

Si tratta di una questione cruciale nell'affrontare la questioni di *explainability* e *fairness* algoritmica, ma anche nell'evidenziarne il sottile, ma pur importante confine che differenzia i due concetti.

L'apprendimento statistico che deriva dalla potenza dei *big data* non può essere considerato come chiave euristica da sola sufficiente alla comprensione della realtà fenomenica e sociale: contrariamente a quanto a suo tempo sostenuto da Anderson[45], prescindere da ogni nesso causale ed accontentarsi che le correlazioni funzionino, significherebbe accettazione fideistica della verità algoritmica: nessun perché di fronte all'oggettività dei dati. Accettazione in cui è ancora possibile incorrere, nonostante il fatto che in tutti i campi della scienza, l'assunto per cui i dati possano "parlare da soli", liberi da pregiudizi umani, posizionamenti o inquadramenti predeterminati, sia stato già efficacemente decostruito[46]. Il dato non esiste come *datum*: è narrazione e rappresentazione e in quanto tale frutto di un'attività umana di interpretazione, giudizio e decisione[47].

Invero, nonostante ad oggi vi sia una sempre maggiore consapevolezza degli effetti discriminatori, attuali e potenziali, in particolare dell'*Algorithmic decision making* sulla società[48], si è lontani dall'elaborazione di una teoria generale e comprensiva, che gettando un ponte semantico tra le scienze informatiche, sociali e giuridiche, possa colmare il gap concettuale attualmente identificato come *bias*[49] nella scienza dei dati. Questo, a sua volta, rende difficile concettualizzare e dare contenuto, dal punto di vista normativo, al

---

e matematiche, ma anche scienze umane. Cfr. C. Anderson, *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, Wired, 23 June 2008, https://www.wired.com/2008/06/pb-theory/.

[44] È significativo che gli esempi utilizzati per dimostrare l'efficacia di questo approccio provengano spesso dagli studi di marketing: esso infatti si dimostra molto efficace nell'individuare correlazioni tra gli acquisti, J. Dyche*, Big data 'Eurekas!' don't just happen*, in *Harvard Business Review Blog*, 20 November 2012, http://blogs.hbr.org/cs/2012/11/eureka_doesnt_just_ happen.html.

[45] C. Anderson, T*he End of Theory: The Data Deluge Makes the Scientific Method Obsolete,* cit.

[46] Molti sono gli studi che hanno messo in evidenza le limitazioni intrinseche ad un approccio esclusivamente basato su inferenza a partire dai dati, e come viceversa rimanga necessario avere una spiegazione del perché la macchina dia quella specifica previsione. Si veda almeno R. Kitchin, *Big data and human geography: Opportunities, challenges and risks. Dialogues in Human Geography*, No. 3, 2013, 262–267; R. Kitchin, *Big data, new epistemologies and paradigm shifts*, in Big *Data & Society,* No. 1, 2014, 1-12; K. Crawford, *The hidden biases of big data*, in *Harvard Business Review Blog*, 2013, http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/; D. Boyd, K. Crawford, *Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon*, in *Information, Communication & Society*, No. 5, 2012, 662-679; A. Testolin, M. Piccolini, S. Suweis, *Deep learning systems as complex networks,* in *Journal of Complex Networks,* No. 1, 2020.

[47] Si veda L. Gitelman, '*Raw Data' is an Oxymoron, Cambridge*, MIT Press, 2013.

[48] Sul tema si veda *ex multis:* J. Van Dijck, *Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. surveillance and society,* No. 12, 2014, 197-208. J. Yu, N. Couldry, *Education as a domain of natural data extraction: analysing corporate discourse about educational tracking*, in *Information, Communication & Society*, No. 1, 2022.

[49] Nel campo dell'IA, il termine *bias* tende a riferirsi generalmente agli effetti di un sistema informatico, che discrimina ingiustamente negando un'opportunità, un bene o assegna un risultato indesiderato a un individuo o a un gruppo per ragioni inappropriate cfr. B. Friedman, H. Nissenbaum. *Biasicomputersystems*, in *ACMTrans.Inf.Syst.*, No. 3, 1996, 330–347.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

concetto che va sotto il nome di *fairness* algoritmica[50], questione strettamente connessa, anche nel linguaggio del legislatore, all'idea di *algorithmic explainability,*

Il parametro della c.d. *algorithmic explainability* è infatti uno dei criteri principali che ispira gli attuali sforzi normativi europei per affrontare le questioni etiche nella regolamentazione dell'IA. Ciò significa che le soluzioni passano in primis attraverso il tentativo di fare luce nella scatola nera, indicando con la trasparenza, declinata in vari modi, la *condicio sine qua non* di garanzia e di monitoraggio della *fairness* algoritmica. Il concetto, tuttavia, necessita di disambiguazioni se si vuole scongiurare il rischio di ancorarlo all'idea di un'utopica oggettività raggiungibile dai sistemi di apprendimento artificiale.

Phillips e Przybocki hanno delineato quattro principi volti a dare contenuto al concetto di *explainability* che si aggiungono a quelli classici di trasparenza, fiducia e correttezza. Tuttavia, anche gli ulteriori principi enucleati, quali comprensibilità e spiegabilità, intesi come la facoltà di "ricevere informazioni significative sulla logica utilizzata" dalla macchina per fornire il risultato, continuano a basarsi sul presupposto che gli algoritmi abbiano una logica, in senso deterministico-matematico, ovvero che vi sia una connessione ripercorribile e verificabile di induzioni correttamente svolte tra principi generali ed applicazioni, basate sul principio di causalità[51]. Abbiamo già avuto modo di mettere in luce, tuttavia, come gli algoritmi di *machine learning* operino (almeno per ora) secondo una logica deduttiva non causale: il sistema "impara" a partire da dati che sono sovente raccolti dalla rete Internet (e che non risultano dunque verificabili integralmente)[52] e si evolve autonomamente, con la conseguenza che anche l'integrale rivelazione del codice sorgente potrebbe non determinare la piena comprensibilità del modo di operare della macchina, che spesso resta ignota agli stessi programmatori[53].

Circostanza, questa, che rende molto difficile, se non impossibile, cogliere l'origine dei potenziali *bias* algoritmici e quindi le modalità attraverso cui intervenire per ottenere un algoritmo conforme ai risultati desiderati. Questo è sicuramente il caso dei cosiddetti "pregiudizi emergenti" (*emergent biases*)[54], che nascosti dietro la neutralità degli iniziali parametri di addestramento, sorgono durante i successivi processi di apprendimento, rivelandosi come distorsioni estremamente difficili, se non impossibili, da prevedere in

[50] Con il termine *fairness* reso con "equità algoritmica" si allude alla capacità degli algoritmi di ridurre e rimuovere le discriminazioni rispetto a determinati attributi sensibili (pensiamo ad esempio ai casi in cui l'algoritmo decide di non concedere un prestito sulla base dell'etnia). Si veda in merito: M. Galeotti *Discriminazione e algoritmi. Incontri e scontri tra diverse idee di fairness*, in *The Lab's Quarterly*, No. 4, 2018, 96.

[51] Così A. Simoncini, S. Suweis, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Rivista di filosofia del diritto*, in *Journal of Legal Philosophy*, No.1, 2019, 87-106.

[52] Si parla a proposito di "*biases* di misurazione". Per una ricostruzione delle categorie di *bias* emerse in letteratura nel campo dell'IA, si veda *idem*, 96.

[53] *Ibidem*; D. Messina, *La proposta di regolamento europeo in materia di Intelligenza Artificiale: Verso una "discutibile" tutela individuale di tipo consumer-centric nella società dominata dal "pensiero artificiale"*, in *Media Laws*, No. 2, 2020, 196-231.

[54] Sul tema si veda F.Z. Borgesius, *Discrimination, Artificial Intelligence and Algorithmic Decision-Making, Anti-discrimination Department of the Council of Europe*, Strasbourg, 2018. Per una dettagliata analisi del fenomeno della *proxy discrimination* e sulle sue molteplici sfumature applicative cfr. A.E.R. Prince, D. Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data,* in *105 Iowa L. Rev*., No. 3, 2020, 1257 ss..

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

fase di progettazione. Le cosiddette *proxy discriminations*[55] che ne derivano, sorgono quali conseguenze involontarie di correlazioni statistiche apofeniche, a loro volta soggette a meccanismi di amplificazione circolare che vanno sotto il nome di *feedback loops*[56].

Pertanto, già la descritta forma di opacità degli algoritmi mette in discussione, dal punto di vista tecnico, l'idea che l'equità possa essere raggiunta esclusivamente facendo luce sui processi matematici e sull'iniziale correttezza dei parametri, che per quanto in apparenza "neutrali", non sono sufficienti a scongiurare *bias* che emergono nel corso dell'apprendimento. Ma non è questo l'unico aspetto a venire in rilievo e che vale la pena di sottolineare. Si potrebbe infatti obiettare che questa declinazione di opacità (c.d. *intrinsic opacity*[57]) che si oppone all'*explainability* e i *bias* che ne derivano, sia suscettibile di rilevazione attraverso meccanismi di supervisione e monitoraggio controfattuale[58]. Non può però dimenticarsi che è possibile far emergere i descritti *biases* nella misura in cui si disponga di un bagaglio etico-normativo condiviso, tale, da poter considerare le correlazioni come errate e/o inaccettabili in relazione agli obiettivi attesi. Se la scelta e la condivisione dei criteri assiologici alla base della decisione algoritmica è fondamentale per immaginare un algoritmo "*fair*"[59] il quesito da porsi non può non investire il parametro etico-normativo a cui fare riferimento per effettuare questa scelta. La questione sembra a prima vista esulare dalle preoccupazioni di parte di quella dottrina che descrive la *fairness* come un tentativo di cancellare le discriminazioni (*debiasing the algorithm*) a partire dall' assunto implicito, discusso in apertura, per cui metodi di raccolta dati (test, esami, analisi a campione) producano una descrizione accurata della realtà, prospettiva sintetizzata con l'acronimo WYSIWYG, che sta per *What You See Is What You Get*[60]. I metodi basati sul WYSIWYG partono quindi dall'idea che esista una classificazione "giusta" degli individui, ed il compito di un algoritmo *fair* sia quello di ricostruire questa classificazione eliminando le discriminazioni individuali basate sugli attributi protetti. Tuttavia l'incompatibilità di criteri che emerge già tra le teorie che adottano questa prospettiva[61] è da sola sufficiente a metterne in luce non solo il *bias*

[55] P. Hacker, *Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against Algorithmic Discrimination Under EU Law*, in *Common Market Law Review,* No. 55, 2018, 1146 ss..

[56] In merito, si veda M. Airoldi, *The spectrum of the algorithm and the social sciences. Critical perspectives on intelligent machines and automation of polis inequalities*, in *Polis (Italy)*, No. XXXIV, 2020, 111-128; M. Airoldi, D. Gambetta, *On the myth of algorithmic neutrality*, in *The Lab's Quarterly,* No. 3, 2018, 25-46. D. Sumpter, *Outnumbered: From Facebook and Google to Fake News and Filter-Bubbles - the Algorithms that Control Our Lives,* Bloomsbury Publishing, 2018.

[57] Per una classificazione tripartita delle diverse declinazioni di opacità algoritmica si veda P. Zuddas che distingue tra opacità tecnica, «opacità tecnica, opacità intrinseca e opacità giuridica» degli algoritmi. In P. Zuddas, *Brevi note sulla trasparenza algoritmica*, in *Amministrazione in Cammino*, Fasc. *2, Diritto di internet - osservatorio sulla discriminazione telematica,* 2020, 1-17.

[58] Il problema non sembra infatti sfuggire alla Commissione Europea che nel Libro Bianco sull'intelligenza artificiale ricorda la necessità di meccanismi di valutazione post-attuazione e un monitoraggio continuo, cfr.*, White paper on Artificial Intelligence - A European approach to excellence and trust,* Brussels, 19.2.2020. È chiara quindi la necessità di una postura etico-normativa chiara e condivisa che possa fungere da parametro controfattuale nella valutazione dell'operato del sistema.

[59] Così S.A. Friedler, C.E. Scheidegger, S. Venkatasubramanian, *On the (im)possibility of fairness*, in *arXiv.org:1609.07236*, 2016.

[60] Si veda in merito M. Galeotti, *Discriminazione e algoritmi. Incontri e scontri tra diverse idee di fairness. Gli algoritmi come costruzione sociale*, cit., nota 40.

[61] Le modellizzazioni della *fairness* possono anche essere incompatibili tra loro e basarsi su assunti che provengono da ipotesi di fondo diverse: se infatti pensiamo all'assenza di discriminazione come una

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

epistemologico, ma anche ciò che precisamente manca alla dimensione dell'*explainability*: una riflessione condivisa ed eventualmente sindacabile, delle differenti visioni di mondo implicite al concetto di *fairness*. Visioni, solo successivamente incorporate nell'algoritmo attraverso la codifica specifica allo strumento matematico utilizzato. In altri termini, a restare in ombra, è la necessità di un substrato etico-normativo condiviso per valutare le modalità e il carattere democratico o meno della decisione (umana) che informa l'algoritmo, e che fornisce i criteri in base ai quali valutare l'algoritmo "giusto" nel merito, oltre che corretto nel procedimento, spiegabile e, dunque, "legittimo". La possibilità di risolvere discriminazioni e disparità di trattamento presenti nella società attraverso una programmazione il più possibile "*unbiased"* è ipotizzabile infatti a condizione che si disponga di categorie interpretative e dunque di un substrato valoriale condiviso, volto a definire la postura etico-normativa dell'algoritmo che si vuole *fair.* Quando si di discute di *fairness* algoritmica se da un lato non può prescindersi da una riflessione più ampia, che renda esplicito il portato etico-normativo dei parametri di cui si chiede "spiegabilità", dall'altro necessita chiarezza circa le opzioni e le valutazioni alternative concretamente e idealmente disponibili. *Explainability* e *fairness* algoritmica vengono, dunque, in rilievo come due concetti strettamente connessi ma non completamente sovrapponibili. Circostanza questa, che non può sfuggire se si presta attenzione a quelle che sembrano atteggiarsi a forme di "distorsione fisiologica" dell'oggettività algoritmica, probabilmente più difficili da individuare, ma non per questo meno rilevanti: si tratta delle categorie ermeneutiche e degli stereotipi cognitivi di cui la macchina, non diversamente dall'agente umano, non può prescindere nel corso dell'apprendimento. I potenziali meccanismi distorsivi derivanti dalla semantica culturalmente e socialmente "situata" dell'algoritmo, potrebbero essere ricondotti a quelli che nel campo dell'IA vengono definiti come *pre-existing biases*[62].

Tuttavia, in considerazione del fatto che è la codificazione sociale e culturale immanente ad un dato contesto a venire in rilievo in questo caso, sulla scorta delle suggestioni di Airoldi[63] potremmo utilmente identificare questa categoria di "distorsioni" come *habitus biases.* Con il termine vogliamo riferirci, infatti, a tutti quei casi in cui non è la correttezza dei codici matematici, né le potenziali distorsioni "precedenti" (*datamining*) o "emergenti" a venire in discussione, ma la "condivisibilità" *ab origine* dei codici sociali e culturali: i valori e disvalori e, ancor prima, le categorie e gli stereotipi

situazione di uguaglianza delle probabilità o delle opportunità avremo una formalizzazione di un algoritmo "*fair*" molto diversa da quella che avremmo se invece pensassimo alla *fairness* come una condizione di parità statistica tra gruppi. Ad es. i metodi WAE acronimo per *We're All Equal.* che hanno come focus quello di evitare discriminazione di gruppo partendo da un'idea di parità statistica tra gruppi, prestano il fianco a critiche in relazione alle discriminazioni individuali che i sistemi così formalizzati potrebbero potrebbe produrre. Cfr. *idem*, 87.

[62] Cfr. A. Simoncini, et al., *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale*, cit., 96; D. Messina, *La proposta di regolamento europeo in materia di Intelligenza Artificiale: Verso una "discutibile" tutela individuale di tipo consumer-centric nella società dominata dal "pensiero artificiale*, cit.

[63] Il concetto di machine habitus è stato elaborato da M. Airoldi, *The machine habitus. Towards a sociology of algorithms*. John Wiley & Sons 2021. Il riferimento è alla nota teoria di Pierre Bourdieu per cui l'habitus non costituisce tanto (o soltanto) il sapere, ma riguarda il rapporto col sapere; un «sistema di disposizioni durabili e trasferibili, strutture strutturate, predisposte a funzionare come strutture strutturanti» per la maggior parte di natura inconscia. Cfr. P. Bourdieu*, Social Space and Symbolic Power,* in *Sociological Theory,* No. 1, 1989, 14-25.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

cognitivi – parte dei processi di apprendimento umano – che parimenti informano, i processi di formalizzazione e apprendimento dell'algoritmo. Se le macchine sono il frutto della creazione umana, e gli esseri umani possiedono diversi sistemi di valori, principi e credenze, socialmente "situati" avviene, in maniera pressoché inevitabile, che tale bagaglio esperienziale, e le relative categorie cognitive, vengano trasferite – intenzionalmente o meno – all'interno di tali sistemi, con la conseguenza di renderli *a priori* non neutrali perché "intrisi", sin dalla loro progettazione, della specifica "visione del mondo" dei loro programmatori e degli individui con cui interagiscono. In altri termini ne vestiranno l'*habitus* socio-culturale[64].

Del resto, in assenza di una cornice interpretativa non sarebbe possibile attribuire alcun significato ai dati né alle loro correlazioni, non riconoscere questo significa semplicemente evitare di esplicitare qual è la propria cornice di riferimento, nascondendola sotto una pretesa di "neutralità" o "oggettività". Come afferma Vis «Raccontiamo storie attraverso i dati ed essenzialmente si tratta delle storie che abbiamo intenzione di raccontare»[65].

Queste considerazioni avvalorano la discutibilità, sul piano epistemologico, dell'idea stessa di ottenere un algoritmo privo di *bias* (e dunque *fair*) fondata sull'assunto implicito che l'algoritmo debba semplicemente avvicinarsi il più possibile a un contesto di selezione "giusto", che esiste "là fuori", indipendentemente dall'algoritmo stesso, e che quest'ultimo possa darne una riproduzione oggettiva e fedele, scevra da pregiudizi e categorie interpretative proprie dell'apprendimento umano.

Abbiamo visto, infatti, come gli assiomi iniziali di un modello condizionino necessariamente i passaggi successivi, ma soprattutto come la realtà su cui gli algoritmi operano, rappresentata e interpretata attraverso i dati di cui si nutrono, non può darsi indipendentemente dall'algoritmo. L'algoritmo, attraverso il linguaggio che gli è proprio, contribuisce a creare e a "re-informare" la realtà in cui opera attraverso un rapporto sinergico di reciproca influenza e riscrittura.

A ciò si aggiunge che la progressiva autonomia dei sistemi di intelligenza generativa vede l'interazione di categorie cognitive, stereotipi e pregiudizi umani rielaborati secondo la logica non abduttiva dell'apprendimento artificiale. Questo sembra dar vita ad una forma di conoscenza nuova e diversa: l'"apofenia" dell'IA sembra rendere inconferente lo iato che esiste tra significato e significante, e che rappresenta un *discrimen* fondamentale nei processi di apprendimento umano, circostanza questa che probabilmente imporrà di riflettere sull'idoneità delle categorie ermeneutiche finora elaborate sul modello dell'intelligenza umana a cogliere e spiegare, per analogia, l'apprendimento artificiale.

La spiegabilità dei criteri di natura proattiva o reattiva[66] che sia, allo stato dell'arte, potrebbe essere da sola insufficiente a fungere da unico presidio dell'*agency* umana. Questo non solo, qualora non sia chiara e democraticamente condivisa la postura etico-normativa che si vuole informi il sistema ed in base alla quale ne viene valutata l'equità

[64] *Ibidem*.

[65] F. Vis*, A critical reflection on big data: considering APIs, researchers and tools as data makers. First Monday, 10*. 2013, http://firstmonday.org/ojs/index.php/fm/article/view/4878/3755.

[66] Si parla di approccio, *ex ante*, qualificabile come "proattivo", nei casi in cui l'algoritmo venga reso "leggibile" già in fase di programmazione; e un approccio, viceversa, *ex post*, qualificabile come "reattivo", che si collega al diritto di accedere al codice sorgente del software, al fine di comprenderne la logica funzionale. Su quest'approccio si veda G. Resta*, Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, in *Politica del diritto*, No. 2. 2019, 223.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

e la sua accettabilità nell'ordine sociale/giuridico, ma anche quando venga meno la consapevolezza della disponibilità e della negoziabilità di alternative possibili.

I «sistemi algocratici»[67], infatti, a differenza di quelli burocratici, strutturano il campo delle azioni possibili senza bisogno che gli agenti interiorizzino il rispetto per regole e leggi, né vi siano indotti dalla cognizione di punizioni: la loro azione è controllata dando forma all'ambiente in cui si svolge, e facendo in modo che siano presenti solo alternative programmate. Se così è, e se la verità algoritmica è un esito possibile tra più opzioni, la trasparenza non può essere da sola sufficiente a porsi come antidoto all'algocrazia. Un riequilibrio in senso democratico non può che passare dalla consapevolezza della "negozialibilità" della verità algoritmica, corredata da un 'effettiva libertà di scelta e azione. Equità algoritmica significa allora, prima di tutto, effettiva possibilità per l'agente umano di contribuire attivamente alla costruzione dell'idea di mondo proposta dall'IA.

Un utile esempio della differenza che intercorre tra il binomio equità procedurale – trasparenza ed il binomio equità sostanziale – *fairness*, può essere tratto dalle nuove disposizioni relative alla disciplina dei Sistemi di Raccomandazione (RS)[68] previste dal regolamento europeo Digital Services Act (DSA)[69] di recente approvazione. Gli RS sono il motore «neo-intermediazione algoritmica»[70] dell'informazione, ragion per cui l'art. 27 del citato regolamento prevede che le piattaforme debbano esplicitare i parametri che informano tali sistemi e l'eventuale disponibilità di criteri alternativi.

Per quanto si tratti di un passo fondamentale in materia di trasparenza, sembra ancora mancare qualcosa: nulla è detto su quali debbano effettivamente essere le opzioni messe a disposizione dell'utente o sul modo in cui queste debbano perseguire obiettivi di interesse pubblico quali eterogeneità delle fonti e il pluralismo informativo. Soprattutto, non può sfuggire il fatto che la formalizzazione di tali sistemi è rimessa ancora e unicamente alle stesse piattaforme in cui gli RS operano[71]. Non è previsto infatti alcun obbligo per le piattaforme di consentire la scelta tra algoritmi di raccomandazione implementate da terze parti (ed esempio tramite sistemi *middleware*[72]) per cui è

[67] Il concetto di "algocrazia" è stato teorizzato per la prima volta da A. Aneesh, docente di sociologia all'Università del Wisconsin-Milwaukee e la nascita della voce *algocracy* viene fatta risalire al 2006, data in cui viene pubblicato il suo libro *Virtual Migration*. «Con il termine algocrazia viene descritto un ambiente digitale di rete in cui il potere viene esercitato in modo sempre più profondo dagli algoritmi, cioè i programmi informatici che sono alla base delle piattaforme mediatiche, i quali rendono possibili alcune forme di interazione e di organizzazione e ne ostacolano altre». A. Delfanti, A. Arvidsson, *Introduzione ai media digitali*, Il Mulino, 2013, 23. Si veda anche F. Gallo, *Democrazia 2.0. La Costituzione, i cittadini e la partecipazione*, Treccani.it, 22 gennaio 2020.

[68] Con la locuzione si fa riferimento a strumenti e tecniche software basati sui dati che forniscono suggerimenti circa le informazioni e gli elementi che possono essere utili ad un utente. Per un'ampia trattazione si veda, M. Hildebrandt, S. Gutwirth, *Profiling the European citizen: cross-disciplinary perspectives,* Springer, 2008; F. Ricci, L. Rokach, B. Shapira, *Recommender systems: Introduction and challenges,* in F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook,* Springer, 2015, 1-34.

[69] Regolamento Ue 2022/2065 (Dsa).

[70] I. de Vivo, *The "neo-intermediation" of large on-line platforms: Perspectives of analysis of the "state of health" of the digital information ecosystem,* in *Communications,* No. 3, 2023, 420-439.

[71] Per tali rilievi critici si veda: N. Helberger, M. Van Drunen, S. Vrijenhoek, J. Möller, *Regulation of news recommenders in the Digital Services Act: Empowering David against the very large online Goliath,* in *Internet Policy Review*, 2021; M. Hildebrandt, S. Gutwirth, *Profiling the European citizen,* cit.

[72] Il *middleware* è software di raccordo tra due applicazioni per trasmettere dati da una all'altra. Nello specifico un «elemento architetturale di un sistema informativo che introduce un livello di

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

prevedibile che questi continueranno ad interpretare il sociale con le stesse categorie cognitive, gli stessi *biases* (intenzionali o meno) e gli stessi obiettivi strategici – presumibilmente orientati alla massimizzazione del profitto – delle piattaforme in cui operano.

A ben vedere, ciò di cui un tale approccio alla trasparenza sembra mancare è l'approfondimento della dimensione etica che il concetto di *fairness* algoritmica presuppone, ossia, in questo caso, la riflessione su quali siano che le condizioni minime a tutela della libertà epistemica sottesa ad una costruzione, che possa dirsi autodeterminata, della propria sfera decisionale e cognitiva.

Le piattaforme, infatti, prima ancora di vincolare l'individuo al modo in cui è "visto", "profilato" e processato algoritmicamente, ne costruiscono le possibilità identitarie definendo (ampliandolo o riducendolo) lo spazio semantico individuale e collettivo: il linguaggio e le categorie cognitive a disposizione per la propria costruzione identitaria.

Si tratta di una dimensione cruciale da considerare se si vuole garantire "l'equità" del linguaggio artificiale: «la libertà epistemica alla base della costruzione moralmente libera dell'identità, è infatti precondizione per garantire la resilienza di quell'autosovranità necessaria e insopprimibile a fungere da argine paradigmi egemonici 'normalizzanti' umani o algoritmici che siano». Per questo motivo, un indirizzo etico-normativo a livello di progettazione non solo deve esistere – se il «codice è legge»[73], il suo contenuto gli obiettivi e le migliori modalità per raggiungerli, non possono essere lasciati alla definizione unilaterale dei loro programmatori[74] e delle piattaforme in cui operano – ma chiara deve esserne la natura socialmente storicamente situata, se si vuole scongiurare forme di egemonia e condizionamento ideologico – e prima ancora cognitivo – nascosti dietro l'utopica oggettività della *praxis* algoritmica. Si tratta in altri termini di garantire la libertà morale di immaginare alternative possibili.

È una sfida alla quale non è possibile sottrarsi se si vuole che i ritmi vertiginosi di sviluppo di dei sistemi di IA generativa non si traducano in contestuale e vertiginoso depauperamento dell'autonomia e dell'*agency* umana, ma vadano ad affiancarla con un bagaglio cognitivo nuovo e diverso e attraverso un tipo di *agency* parallela e non sovrapponibile. La tutela di una sfera incomprimibile a garanzia della libertà epistemica dell'agente umano, è dunque primo presupposto per evitare che il rapporto simbiotico tra *agency* artificiale e *agency* umana non si trasformi in un depauperamento parassitario a scapito di quest'ultima.

Parallelamente, quella che, al contrario, sembra prospettarsi come una progressiva "autonomia cognitiva artificiale" alla base del crescente "affrancamento decisionale" dei sistemi di IA dal fattore umano ha avuto e, non poteva non avere riflessi, anche nel diritto, aprendo il dibattito circa la necessità di rivedere le tradizionali categorie giuridiche e se

disaccoppiamento tra il nucleo centrale del sistema (che potrebbe essere costituito da un server o da un mainframe) e le sue parti periferiche, svolgendo, di fatto, un ruolo di mediazione tra i dati e le informazioni elaborati a livello centrale e ciò che viene gestito direttamente a livello di interfaccia con l'utente» (https://www.treccani.it/enciclopedia/middleware/) sul tema si veda F. Fukuyama, B. Richman, A. Goel, *How to Save Democracy From Technology. Ending Big Tech's Information Monopoly,* in *Foreign Affairs*, January/February 2021.

[73] Il riferimento è L. Lessig, *Code and other laws of cyberspace, version 2.0*., Basic Books, 2006.

[74] Ad esempio, una decisione di IA basata su criteri di uguaglianza formale o di base funzionerà in modo molto diverso e produrrà risultati molto diversi rispetto a una decisione basata su principi di uguaglianza più profondi.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

del caso crearne di nuove. In particolare oggetto di dibattito, è la possibilità di introdurre, nel *framework* normativo, un inedito *tertium genus*, interposto tra persone fisiche e persone giuridiche: una *fictio iuris* ultronea, necessaria ad inquadrare il peculiare *status* dell'"agente elettronico" nella sua capacità di comunicazione, interazione e interferenza con l'*agency* umana.

3.1. *Verso nuove categorie giuridiche? La via della personalità elettronica*

Per risolvere le inedite sfide giuridiche che l'*AI Act* si trova ad affrontare, come l'urgenza di elaborare adeguati criteri d'imputazione di responsabilità per danni causati da algoritmi, una questione preliminare è infatti proprio la necessità di sciogliere il nodo relativo allo *status* – da riconoscere a tali "sistemi agenti" – ai fini di un corretto inquadramento giuridico, in grado di tener debitamente conto della loro sempre maggiore autonomia decisionale ed indipendenza dal fattore umano.

Il carattere senza precedenti della sfida deriva prima di tutto dal ruolo che tali "sistemi agenti" assumono nella realtà 4.0: il grado di autonomia decisionale sempre maggiore che fa sorgere la spinosa questione etica, prima ancora che giuridica, circa la natura della loro "soggettività".

Il legislatore europeo non ha potuto, dunque, non interrogarsi sull'adeguatezza del tradizionale quadro teorico-giuridico di riferimento, data l'urgenza di sciogliere una serie interrogativi che necessitano di discussione: dall'opportunità o meno di un riconoscimento di autonomia giuridica e, in caso di risposta affermativa, se orientarsi secondo paradigmi già esistenti o crearne di nuovi. La creazione di un inedito *tertium genus,* la c.d. personalità elettronica, così come suggerito dal Parlamento europeo già a partire dalla risoluzione del 16 febbraio 2017 (c.d. Carta della Robotica)[75] permetterebbe di andare oltre la personalità giuridica riconosciuta alle aziende. Ciò consentirebbe la diretta attribuzione all'«agente elettronico»[76] non solo di responsabilità, ma anche di diritti, subordinando l'identità sociale e capacità di agire dell'algoritmo alla sua autonoma capacità di decisione e «comunicazione»[77].

[75] P8TA (2017)0051, *Norme di diritto civile sulla robotica, Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica* (2015/2103(INL). c.d. "Carta della Robotica". La Risoluzione riprende integralmente il rapporto Delvaux, raccomandando all'art. 59 lett. f), l'«Istituzione nel lungo termine di uno status giuridico specifico per i robot in modo che almeno i robot 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 i terzi».

[76] P8TA (2017)0051, *Norme di diritto civile sulla robotica, Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica* (2015/2103(INL). c.d. "Carta della Robotica". La Risoluzione riprende integralmente il rapporto Delvaux, raccomandando all'art. 59 lett. f), l'«Istituzione nel lungo termine di uno status giuridico specifico per i robot in modo che almeno i robot 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 i terzi».

[77] «L'autonomia di un robot può essere definita come la capacità di prendere decisioni e metterle in atto nel mondo esterno, indipendentemente da un controllo o un'influenza esterna (…) tale autonomia è di natura puramente tecnologica e il suo livello dipende dal grado di complessità con cui è stata progettata l'interazione di un robot con l'ambiente.» Cfr. *supra* nota 56.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

Alla luce, tuttavia, delle sfide che un'entificazione in senso personalistico degli algoritmi comporterebbe, quali l'impatto etico dell'eventuale attribuzione alla macchina non solo di capacità cognitive ma anche emotive, si comprende come la personalità elettronica, nonostante il *favor* incontrato in seno al Parlamento Europeo, abbia ancora molta strada da percorrere.

Anche l'ipotesi sostenuta dalla citata *Carta della Robotica* che vorrebbe subordinare l'attribuzione di una «autonomia digitale graduabile», e quindi di una «identità sociale e capacità di agire» dell'algoritmo, alla sua autonoma capacità di decisione e «comunicazione», prescindendo da questioni riguardanti la capacità di «autocoscienza» o i «processi psicologici»[78], sembra ad oggi risultare di difficile applicazione.

Le macchine intelligenti, pur essendo capaci di autonomia decisionale e comportamentale, allo stato dell'arte non possono essere identificate con un'unica definizione che le accomuni tutte indistintamente, né sembrano esserci i criteri per graduare quest'autonomia secondo criteri condivisi. Sono molteplici, infatti, gli aspetti che necessitano di ulteriori indagini quali la loro natura, l'ambiente in cui operano e il tipo di controllo esercitabile su di esso.

Per un corretto approccio legislativo e regolamentare in materia di IA sarà fondamentale, quindi, cercare di comprendere con chiarezza ciò che l'AI può fare, ciò che non può fare e ciò che, nel breve, medio e lungo termine, sarà in grado di fare. Pertanto, la corrente di pensiero a favore dell'autonomia o della almeno parziale soggettività degli "agenti software", "de-antropomorfizzando" il concetto di entità agente dotata di autonomia, necessita senza dubbio di approfondimento.

Pur volendo, infatti, prescindere dal riconoscimento di una qualche "personalità elettronica", al crescere del livello di intelligenza in talune entità robotiche, oggi già capaci di creazioni tecniche, artistiche e di invenzione, si accompagna e si accompagnerà, in ogni caso, il problema del ritenerle (o di chi ritenere) titolari o "autori" (e responsabili) delle proprie creazioni.

Secondo gli attuali *framework* normativi ed in accordo con la giurisprudenza maggioritaria in ambito tanto internazionale che europeo, né l'ente robotico, né chi lo utilizza potrebbero essere riconosciuti titolari dei diritti morali d'autore, che spettano alla sola persona fisica. Emblematico al riguardo è il caso DABUS (acronimo di *Device for the Autonomous Bootstrapping of Unified Sentience*); si tratta dell'intelligenza artificiale connessionista[79] ideata dallo scienziato statunitense Stephen Thaler, la cui vicenda è stata sottoposta contestualmente all'attenzione del potere giudiziario negli Stati Uniti, Regno Unito, Sud Africa, Australia ed Europa, sollevando un dibattito globale sul ruolo dell'intelligenza artificiale nel contesto attuale.

Lo *European Patent Office* (EPO)[80] – nel rigettare le domande di brevetto per due invenzioni attribuibili a DABUS – con due pronunce confermate in appello dal *Legal*

[78] Si veda in merito U. Ruffolo, *Il problema della "personalità elettronica*, in *Journal of Ethics and Legal Technologies*, No.1, 2020, 75-88.

[79] La macchina si compone di due reti neurali artificiali in grado di elaborare informazioni note al fine di elaborare idee innovative. In sintesi, l'algoritmo consente alla macchina di sviluppare soluzioni originali ma anche di valutare in modo "critico" le idee generate dalle interconnessioni del sistema per determinarne le possibilità di successo e il carattere inventivo.

[80] Provvedimenti n. 18275163 e n. 18275174, 27-01-2020. Il testo delle decisioni è disponibile ai seguenti link: https://register.epo.org/application?documentId=E4B63OBI2076498&number=EP18275174&lng

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

Board of Appeal con decisioni pubblicate il 5.7.2022 e il 4.8.2022 (domande EP 18 275 163 e EP 18 275 174) oltre a fornire un'interpretazione rigorosamente letterale della Convenzione sul brevetto europeo e del termine *inventor,* ha sottolineato che la designazione di un inventore non ha lo scopo meramente formale di garantire certezza, ma è funzionale al riconoscimento di una serie di diritti brevettuali che postulano, all'interno del quadro normativo attuale, la necessaria capacità giuridica dell'inventore. Tale approccio è seguito non solo dalla giurisprudenza di diversi stati membri della Convenzione, ma anche da parte di numerosi uffici brevettuali nazionali. Di conseguenza, allo stato attuale, poiché all'intelligenza artificiale non è riconosciuta alcuna capacità giuridica, prerogativa delle persone fisiche e giuridiche, essa non può essere titolare né esercitare alcun diritto previsto dalla convenzione. L'EPO, inoltre, specifica che l'Intelligenza artificiale non potrebbe nemmeno essere accostata ad una persona giuridica – che, al contrario delle persone fisiche, è titolare di diritti sulla base di una *fictio iuris* – in quanto tali finzioni giuridiche vengono espressamente introdotte e regolamentate dai singoli legislatori nazionali e, ad oggi, nessuna legislazione nazionale ha espressamente riconosciuto tale qualità in capo all'intelligenza artificiale.
Per tali motivi, secondo l'EPO non si può fare applicazione dell'art. 60 della Convenzione, invocato dal richiedente. La disposizione regola l'istituto dell'invenzione del dipendente, ma un'intelligenza artificiale, non dotata di capacità giuridica, non può essere considerata come parte di un contratto di lavoro. Infine, conclude l'EPO, l'intelligenza artificiale non può nemmeno essere accostata a soggetti incapaci di esercitare i propri diritti, come i minori e gli inabilitati, in quanto anche tali soggetti sono dotati di capacità giuridica, sebbene non siano in grado di esercitare autonomamente i propri diritti. In ogni caso, anche in tale ipotesi sono i singoli legislatori nazionali a definire attraverso specifiche disposizioni quali soggetti possano esercitare i propri diritti mediante l'interposizione di un terzo.

Similmente, negli Stati Uniti il rigetto da parte dello U.S. PTO (*United States Patent and Trademark Office*) è stato motivato soprattutto sulla base dell'impossibilità di riconoscere diritti in favore di una macchina, in quanto entità priva di capacità giuridica. Peraltro, sulla base di motivazioni analoghe, il 14.02.2022, il *Copyright Office* degli Stati Uniti ha definitivamente rigettato anche la richiesta di tutelare attraverso il copyright l'opera d'arte creata da un'altra macchina, sempre ideata dal Dott. Thaler.

Anche la Corte d'Appello del Regno Unito, il 21.9.2021, ha confermato la precedente decisione della High Court, rigettando entrambe le domande di brevetto perché l'inventore designato non era un essere umano. Tuttavia, la decisione non è stata presa all'unanimità dei Giudici della Corte d'Appello.

In Australia, dopo un primo rigetto delle domande, il 20.07.2021 c'era stato invece un accoglimento giudiziale da parte della Corte australiana, secondo la quale la nozione di "inventor" contenuta nel *Patents Act* non si riferisce esclusivamente alle persone fisiche. Tuttavia, in seguito all'appello, la Corte Federale Australiana ha ribaltato la decisione del primo giudice riallineando la posizione australiana a quella di Regno Unito, Stati Uniti ed Europa.

Ad oggi l'unica voce dissonante è quella dell'Ufficio Sud Africano, il primo al mondo ad aver attribuito ad un soggetto non umano la titolarità di diritti, riconoscendo a DABUS

=en&npl=false; https://register.epo.org/application?documentId=E4B63SD62191498&number=EP1827 5163&lng=en&npl=false.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

la qualità di inventore: il 24.6.2021 le due domande di brevetto sono state infatti ritenute conformi al *Patents Act* del 1978 e pubblicate ufficialmente sul *South African Patent Journal*.

Il caso DABUS ha dunque dimostrato come le posizioni dei diversi Uffici riceventi non siano perfettamente allineate e, soprattutto, che le interpretazioni delle disposizioni normative in materia di innovazione arranchino a tenere il passo con l'evoluzione tecnologica.

Al di là della questione aperta dell'attribuzione di diritti, consequenziale al riconoscimento di una qualche autonomia giuridica all'agente software, altra questione cruciale che necessita di immediata risposta è il difficile, a volte impossibile, collegamento tra attività delle macchine e responsabilità umana.

Il problema nasce dalla natura multidimensionale e multistrato del processo algoritmico e quindi alla molteplicità di attori che assumono posizioni diverse in tale processo: dalla proprietà dei dati, alla titolarità di software, all' eventuale presenza di contratti di fornitura di servizi (per cui il titolare dello strumento può essere diverso dal titolare del servizio). Tuttavia, l'opacità, la vulnerabilità, la capacità di modifica mediante aggiornamenti, ma soprattutto l'autoapprendimento e la potenziale autonomia dei sistemi di AI, rappresentano la sfida maggiore per l'efficacia dei quadri normativi dell'Unione e nazionali in materia di responsabilità.

La questione di fondo da sciogliere infatti è la seguente: la responsabilità è ascrivibile sempre e solo ad un soggetto umano oppure, in caso di danni a terzi, è necessario riconoscere-autonomia decisionale agli algoritmi?

Un'eventuale attribuzione di personalità giuridica, secondo il modello attualmente previsto per le aziende, seppur non scevra da problematiche, risponderebbe quantomeno all'avvertita esigenza di imputabilità autonoma (ed eventuale corresponsabilità) del sistema agente. Nella pratica, come detto, potrebbe essere molto difficile o addirittura impossibile ricondurre specifiche azioni dannose dei sistemi di AI a uno specifico *input* umano o a decisioni adottate in fase di progettazione. Preoccupazione questa che emerge chiaramente nella citata Risoluzione del parlamento europeo c.d. *Carta della Robotica* in cui si afferma che: «[…] nell'ipotesi in cui un robot possa prendere decisioni autonome, le norme tradizionali non sono sufficienti per attivare la responsabilità per i danni causati da un robot, in quanto non consentirebbero di determinare qual è il soggetto cui incombe la responsabilità del risarcimento né di esigere da tale soggetto la riparazione dei danni causati»[81].

Malgrado ciò, considerando le difficoltà a livello di armonizzazione che l'esplicita attribuzione della responsabilità in relazione ai sistemi di IA comporterebbe, alla luce delle diverse condizioni economiche, giuridiche e sociali di ciascun Stato membro[82], si è

[81] *Carta della Robotica*, v. *supra* nota 80.

[82] Stando al considerando (*d*.) della Risoluzione P9_TA(2020)*0276*, *Regime di responsabilità civile per l'intelligenza artificiale Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale*, «qualsiasi futura legislazione dell'Unione avente come obiettivo l'esplicita attribuzione della responsabilità in relazione ai sistemi di intelligenza artificiale (IA) dovrebbe essere preceduta da un'analisi e dalla consultazione con gli Stati membri riguardo alla conformità dell'atto legislativo proposto alle condizioni economiche, giuridiche e sociali». Cfr. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_IT.html

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

preferito "aggirare l'ostacolo"[83] facendo ricorso ai criteri di attribuzione di responsabilità civile già esistenti e comunemente accettati, (in particolare richiamando la responsabilità per danno da prodotti difettosi), chiamando a rispondere le varie persone nella catena del valore che creano, il sistema di AI e ne eseguono la manutenzione ed il controllo[84].

Attraverso il ricorso all'ormai popolare concetto di "rischio", quale *proxy* nel bilanciamento dei diritti e degli interessi coinvolti, si prevedono infatti, come vedremo, criteri d'imputazione che vanno dalla responsabilità oggettiva (per i sistemi c.d. ad alto rischio) alla responsabilità per colpa.

In Italia, la dottrina aveva già indicato diverse possibili strade per rispondere al quesito prendendo in considerazione le forme di responsabilità oggettiva previste dall'ordinamento italiano (es. responsabilità per le cose pericolose, art. 2050 c.c.; per danni cagionati da cose in custodia, art. 2051 c.c.; per animali, art. 2052 c.c.; per motoveicoli, art. 2054 c.c.; responsabilità del produttore, art. 114, d.lgs. 206/2005), o ancora facendo ricorso al frame delle responsabilità vicarie (es. responsabilità di genitori, tutori e maestri per i danni cagionati da fatti illeciti dei minori e degli allievi (art. 2048 c.c. *culpa in educando*) o alla responsabilità per l'attività di dipendenti, commessi o domestici (art. 2049 c.c. *culpa in vigilando*).

Ad ogni modo, qualsiasi siano le soluzioni accolte, l'obiettivo ultimo di qualsiasi quadro in materia di responsabilità dovrebbe essere quello di garantire la certezza del diritto per tutte le parti, che si tratti del produttore, dell'operatore, della persona interessata o di terzi. Un quadro giuridico in materia di responsabilità civile orientato al futuro deve infatti infondere fiducia nella sicurezza, nell'affidabilità e nella coerenza di prodotti e servizi, al fine di trovare un equilibrio tra l'efficace ed equa tutela delle potenziali vittime di danni o pregiudizi. Allo stesso tempo, occorre garantire la disponibilità di una sufficiente libertà d'azione per consentire alle imprese, in particolare alle PMI, di sviluppare nuove tecnologie, nuovi prodotti e servizi.

In ragione di ciò, come sottolineato in più occasioni dal Parlamento Europeo, appare quanto mai necessario un quadro giuridico orizzontale e armonizzato, basato su principi comuni, per garantire la certezza giuridica, fissare norme uniformi in tutta l'Unione e tutelare efficacemente i valori europei e i diritti dei cittadini.

Secondo la UE, infatti, per sfruttare in modo efficiente i vantaggi e prevenire possibili usi impropri dei sistemi di AI, nonché evitare la frammentazione normativa nell'Unione, è essenziale primariamente disporre in tutta l'Unione, per tutti i sistemi di AI, di una regolazione uniforme, basata su principi e adeguata alle esigenze future.

3.2. *La strategia europea in materia di responsabilità: il* risk-based approach

[83] Il PE «è del parere che l'opacità, la connettività e l'autonomia dei sistemi di IA potrebbero rendere, nella pratica, molto difficile o addirittura impossibile ricondurre specifiche azioni dannose dei sistemi di IA a uno specifico input umano o a decisioni adottate in fase di progettazione; ricorda che, conformemente a concetti di responsabilità ampiamente accettati, è tuttavia possibile aggirare tale ostacolo considerando responsabili le varie persone nella catena del valore che creano il sistema di IA, ne eseguono la manutenzione o ne controllano i rischi associati»; cfr. *idem,* punto 7.

[84] Cfr. punto 6, al punto 8 ritiene che la direttiva sulla responsabilità, per danno da prodotti difettosi, possa essere un mezzo efficace per ottenere un risarcimento ma che dovrebbe ciononostante essere rivista per adattarla al mondo digitale e per affrontare le sfide poste dalle tecnologie digitali emergenti, valutando ad esempio l'inversione delle norme che disciplinano l'onere della prova per i danni causati dalle tecnologie digitali emergenti in casi chiaramente definiti e previa un'adeguata valutazione.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

Il concetto di "responsabilità" svolge un duplice ruolo importante nella nostra vita quotidiana: da un lato, garantisce che una persona vittima di un danno o pregiudizio abbia il diritto di chiedere un risarcimento alla parte di cui sia stata dimostrata la responsabilità di tale danno o pregiudizio e di ricevere il risarcimento dalla stessa; dall'altro lato, fornisce incentivi economici alle persone fisiche e giuridiche affinché evitino sin dall'inizio di causare danni o pregiudizi.

In accordo con il parere espresso dal Parlamento Europeo nella citata risoluzione[85], sulla base delle sfide giuridiche che i sistemi di IA pongono per i regimi di responsabilità civile esistenti, è parso ragionevole istituire un regime comune di responsabilità oggettiva[86] (e sistemi di assicurazione obbligatoria) per i sistemi di IA autonomi cosiddetti ad "alto rischio". In questa categoria rientrano ad esempio, i casi in cui l'IA è usata per infrastrutture critiche come i trasporti, l'accesso all'istruzione o la progettazione di software per la gestione dei lavoratori o di servizi pubblici e privati essenziali.

In linea con i regimi di responsabilità oggettiva degli Stati membri, sempre secondo il PE, il regolamento proposto dalla Commissione dovrebbe, altresì, contemplare le violazioni dei diritti fondamentali come il diritto alla vita, alla salute, all'integrità fisica e al patrimonio, fissando gli importi e l'entità del risarcimento nonché i termini di prescrizione.

Il PE non ha mancato poi di sollecitare la Commissione europea ad un'analisi approfondita delle tradizioni giuridiche degli Stati membri in materia di risarcimento per i danni non patrimoniali, al fine di valutare se l'inclusione dei danni non patrimoniali in atti legislativi specifici relativi all'IA sia necessaria e se sia in contrasto con il quadro giuridico dell'Unione in vigore o pregiudichi il diritto nazionale degli Stati membri.

Tutte le attività, i dispositivi o i processi guidati da sistemi di IA che possono provocare danni o pregiudizi, ma che non sono indicati nell'elenco dei sistemi ad alto rischio, anche secondo il parere del PE, dovrebbero invece continuare a essere soggetti a un regime di responsabilità per colpa. In questo caso la persona interessata, comunque, deve poter far valere una presunzione di colpa dell'operatore, su cui cadrebbe l'onere di dimostrare di aver rispettato gli obblighi di diligenza.

Come accennato, la graduazione degli obblighi e i criteri d'imputazione della responsabilità, qui come in altri strumenti normativi emanati dall'UE, fa perno sul concetto chiave di rischio. Si tratta infatti del criterio di bilanciamento essenziale già alla base del GDPR[87], e seppur con una declinazione parzialmente diversa, dal DSA[88].

Tecnicamente parlando, il "rischio" è una combinazione tra la probabilità che si verifichi un determinato pericolo e l'entità delle sue conseguenze e può quindi servire come *proxy* per il processo decisionale, basato sulla previsione di eventi futuri positivi o

[85] Risoluzione PE 2020-0276, https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_IT.html cit. nota 82.
[86] In questo caso, è sufficiente dimostrare danno e connessione causale tra questo e il funzionamento che lo ha generato ed una complessa valutazione dei rischi e della capacità di controllo da parte persona che, in determinate circostanze, è in grado di minimizzare i rischi e affrontare l'impatto negativo dell'intelligenza artificiale, si veda: G. Comandè, *Intelligenza artificiale e responsabilità tra liability e accountability. Il carattere trasformativo dell'IA. Analisi Giuridica dell'Economia,* Il Mulino, No. 1, 2019, 169-188.
[87] Fra tutti AI-Act COM/2021/206 final.
[88] Regolamento Ue 2022/2065 (Dsa).

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

negativi[89]. Ciò avviene principalmente attraverso le pratiche di analisi del rischio, che consistono in un insieme di metodologie, modelli e processi[90]. La regolamentazione del rischio può quindi essere percepita come un tentativo di affrontare l'ascesa di quella che è stata definita da Ulrich Beck la «società del rischio»[91], attraverso un approccio razionale e tecnocratico che favorisca una *governance* più efficiente ed equa, combattendo al contempo contro «l'eccesso di regolamentazione, le regole legalistiche e prescrittive e gli alti costi della regolamentazione»[92]. Piuttosto che limitarsi a stabilire nuovi diritti e garanzie, l'Unione ha cercato di regolare i rischi aumentando la responsabilità del settore pubblico e privato.

Il rischio è diventato infatti un elemento centrale della legislazione europea contemporanea in relazione alle tecnologie digitali e alle sfide che caratterizzano la società algoritmica. Il GDPR e il DSA sono i precedenti esempi dei numerosi strumenti legislativi che l'Unione ha adottato, o prevede di adottare, utilizzando il rischio come criterio di bilanciamento essenziale per promuovere i diritti umani digitali e garantire, al contempo, il pieno sviluppo del mercato unico digitale.

La regolamentazione del rischio ha acquisito uno slancio via via crescente in tutte le democrazie occidentali ed è diventata sempre più utilizzata come strumento normativo per promuovere le politiche dell'Unione in una serie di settori operativi. Il ruolo ultimo del rischio come tecnica di bilanciamento consente di tracciare un collegamento tra tali disposizioni e i contenuti della tradizionale regolamentazione basata sui diritti, alla luce dell'esperienza costituzionale europea caratterizzata dalla ricerca di un equilibrio equo e proporzionato tra i diversi interessi rappresentati dall'Unione. La dignità umana necessita del rischio che è elemento connaturato alla libertà quindi incomprimibile.

Il *fil rouge* alla base delle varie declinazioni nelle politiche dell'Unione è proprio l'obiettivo di contribuire a creare un ambiente digitale che abbracci i valori e i principi costituzionali sanciti dalla Carta dei diritti fondamentali. Come sostenuto da Giovanni De Gregorio e Pietro Dunn[93] la lente del "costituzionalismo digitale" è allora la chiave ermeneutica essenziale per leggere strategia digitale dell'Unione in grado di dare spessore giuridico al bilanciamento tra istanze di innovazione e diritti nello sviluppo dell'IA, tra l'idea del rischio e il suo *pendant* ossia l'*accountability*.

4. *Alcune considerazioni conclusive e i rischi dell'ordine giuridico dell'algoritmo*

Quanto abbiamo detto fino ad ora ci permette di svolgere un passaggio ulteriore sottolineando gli effetti sulle istituzioni democratiche costituzionali dell'attività ordinatrice dell'algoritmo e viceversa delle stesse, attraverso la funzione ordinatrice del diritto, sull'IA; in particolare l'idea che si sta affermando è che il prodotto – ossia l'ordine creato – dall'algoritmo, per il solo fatto che sia creato da una macchina e che semplifichi

[89] R. Gellert, *The Risk-Based Approach to Data Protection,* Oxford University Press, 2020, 27.
[90] L'analisi del rischio comprende due fasi: la prima è la valutazione del rischio, cioè la misurazione del rischio stesso, che rappresenta la componente scientifica e quantitativa; la seconda, cioè la gestione del rischio (*stricto sensu*), è la componente politica e consiste nella fase decisionale.
[91] U. Beck, *La società del rischio*, Carocci, 2000.
[92] M. Macenaite, *The "Riskification" of European Data Protection Law through a Two-fold Shift European Journal of Risk Regulation*, 2017, 506-509.
[93] G. De Gregorio, P. Dunn, *The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age*, in *Common Market Law Review*, No. 2, 2022, 473-500.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

la vita degli essere umani spaventati dall'impossibilità di gestire l'infinito delle possibilità che si aprono loro innanzi, sia assunto come ordine naturale delle cose; sia, insomma, accettato in senso acritico e fideistico con tutto ciò che comporta a livello cognitivo, esperienziale e culturale per gli individui, la cui identità ne è ampiamente condizionata. Un condizionamento che si riflette sulle istituzioni politiche create dagli individui[94].

La straordinaria potenza di calcolo ed il tipo di apprendimento che, in quanto statistico, sembrerebbe non richiedere una reale comprensione dei fenomeni, rischia infatti di riportare il paradigma dei *big data* ad un presunto piano dell'oggettività. In tale piano sarebbero gli stessi dati, senza alcuna pregiudiziale e senza essere condizionati dall'orizzonte di attese dell'osservatore, a dirci del *benchmark*, del modello e della correlazione significativa fra un numero tendenzialmente infinito di variabili. Si tratta di un pregiudizio in cui è facile incorrere, nonostante il fatto che in tutti i campi della scienza l'assunto per cui dati possano "parlare da soli", liberi da pregiudizi umani, posizionamenti o inquadramenti predeterminati, sia stato già efficacemente decostruito[95]. La precondizione per poter efficacemente smascherare l'«hidden curriculum»[96] dei sistemi di *machine learning* consiste nella consapevolezza critica di come dati e modelli statistici plasmino la realtà nel momento stesso in cui la rappresentano; della loro capacità di porre un ordine, quindi, che si impone di fatto all'uomo.

Riconoscere il carattere performativo di dati e algoritmi e la loro natura di costrutti socio-tecnici[97] in grado di plasmare in maniera sinergica la realtà in cui operano costituisce il primo passo allora per dare adeguata risposta agli interrogativi che fanno da sfondo alla strategia europea per una nuova etica digitale, costruendo una direzione analitica in grado di contribuire ad uno sviluppo etico dell'IA nell'ambito di un quadro normativo integrato a livello sovranazionale.

I processi normativi co-producono gli algoritmi che regolano: la cd «algoretica»[98] va, infatti, di pari passo con lo sviluppo della AI. Si parla e si auspica infatti una etica *by design* incorporata nei *software* in grado di indirizzare e rendere trasparenti le scelte valoriali sottese all'attività decisionale degli agenti elettronici. A prescindere dalla spinosa questione dell'eventuale riconoscimento di soggettività giuridica, già a partire dal riconoscimento dell'algoritmo nella sua qualità di «attante»[99], e in quanto tale capace di modificare l'ecosistema in cui opera, è una precondizione per leggere e valutare le sfide

[94] S. Tiribelli, *Identità personale e algoritmo*, cit.

[95] Sul tema si veda *ex multis:* J. Van Dijck, *Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology, i*n *surveillance and society*, 12, 2014, 197-208; J. Yu & N. Couldry. *Education as a domain of natural data extraction: analysing corporate discourse about educational tracking*, in *Information, Communication & Society,* No. 1, 2022, 127-144.

[96] Cfr. P. Mertala, *Data (il)literacy education as a hidden curriculum of the datafication of education* in *Journal of Media Literacy Education*, No. 3, 2020, 30-42.

[97] Cfr. D. Beer, *The Social Power of Algorithms*, in *"Information, Communication & Society",* No. 1, 2017, 1-13; B. Aragona, C. Felaco, M. Marino, *The Politics of Big Data Assemblages,* in *Partecipazione e conflitto,* No. 2, 2018, 448-471.

[98] P. Benanti, *Oracoli. Tra algoretica e algocrazia*, Luca Sossella Editore, 2018.

[99] B. Latour, *Una sociologia senza oggetto? Note sull'interoggettività,* in E. Landowski, G. Marrone (a cura di), *La società degli oggetti Problemi di interoggetività,* Meltemi, 2002, 203-232; V. Orig, *Une sociologie sans objet? Note théorique sur l'interobjectivité,* in *Sociologie du travail,* 587-607, 1994, republication in *Octave Debary Objets et mémoires*, MSH-Presses de l'Université Laval, 2007, 38-57.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

sottese all'*AI Act*[100] che come abbiamo avuto modo di analizzare persegue l'obiettivo principe di dettare regole armonizzate sull'intelligenza artificiale antropocentrica.

Se la capacità ordinatrice dell'algoritmo dovesse occupare tutto il nostro spazio relazionale, si costruirebbe un ordine che entrerebbe necessariamente in conflitto con quello umano fondato sulle regole giuridiche scelte dalla politica legittimata dal costituzionalismo democratico. Il comportamento umano tenderebbe a subire questo ordine (per comodità, per pigrizia, per paura di non avere un ordine…) con una proiezione sull'azione dell'individuo nella democrazia liberale che di fatto sarebbe determinata dalla *box* dell'IA.

Esistono tre strumenti complementari per garantire il controllo umano *sull'*algoritmo ed evitare il controllo umano "da parte" dell'algoritmo e, in particolare, l'indebolimento della funzione ordinatrice del diritto fondata su scelte valoriali e politiche degli individui sottoposti alle regole stesse.

La prima strada è quella che pretende la maggiore trasparenza possibile nel funzionamento dell'algoritmo che, come abbiamo visto, si fonda sui principi di trasparenza, conoscibilità ed eticità: agendo sulla formula e sull'output dell'IA.

La seconda è, come ha analizzato Irti per il mercato, escludere l'IA da alcuni ambiti, insomma, "affamarla". Questo non vuol dire rinunciarvi, ma farla rimanere davvero un mero mezzo, parzializzarla, dividerla come il costituzionalismo ci ha insegnato a dividere ogni potere (pubblico e privato) in grado di ordinare la vita umana. Non occuparsi, dunque, esclusivamente di come far funzionare bene ed eticamente l'IA, ma individuare altresì le categorie di contesti in cui evitare il suo impiego.

La terza infine consiste nella costruzione di una capacità diffusa umana di acquisire consapevolezza su questo processo di intermediazione nelle relazioni umane e di costruzione di un ordine autonomo distinto da quello normativo legittimato dal costituzionalismo democratico. Occorre evitare che l'IA diventi una intelligenza media che si impone come chiave di decodifica del mondo, con un effetto di omogeneizzazione della dimensione relazionale umana con la scusa del fine di mettere ordine all'infinito che si è spalancato davanti alle nostre menti limitate. Possiamo governare l'abbondanza dei dati senza dover rinunciare alle peculiarità umane? Il nostro cervello non concepisce l'infinito e opera sapendo di essere finito laddove ogni notte cancella dalla memoria gli accadimenti che ritiene superflui. È progettato per conferire all'uomo un ordine finito attraverso strumenti finiti per realtà che, per quanto ampie, sono finite.

Non potendo sostituire il proprio cervello limitato con uno illimitato senza perdere la natura umana e le modalità di gestione del proprio mondo che su quella limitatezza sono fondate, l'uomo deve ricondurre l'IA a quello che il cervello umano può governare: un aratro iper-innovativo, ma comunque un aratro, decidendo per quali terreni e quali culture usarlo, cercando di prevedere come esso lavori in tali casi; al tempo stesso l'uomo deve decidere dove non impiegarlo, ossia dove non permettergli di svolgere una funzione ordinatrice che è potenzialmente sostitutiva di quella umana e distruttiva dell'ordine umano fondato sul diritto politico.

Due intelligenze che convivono, lasciando tuttavia a quella umana la potestà di dettare le regole a quella artificiale, non permettendo alla prima di imporre un ordine regolatorio deciso da pochi o, addirittura, dalla macchina stessa; un ordine che rischia di essere

[100] Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) COM/2021/206 final cit.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

assunto dall'essere umano come unico ordine possibile per contenere le paure che ha l'uomo – ma non la macchina – davanti all'infinito relazionale delle possibilità e dei rischi. La nostra limitatezza è dunque la forza di dividere, come per il potere del Re assoluto, del Governo democratico, dei Parlamenti elettivi, del mercato finanche il potere dell'algoritmo.

Sinossi.

Gli esseri umani hanno storicamente sviluppato strumenti, sia tangibili che intangibili, per imporre l'ordine, mitigare i rischi e allineare le azioni con valori assunti a parametro collettivo ella qualificazione dei comportamenti umani e delle loro relazioni. Il diritto, prima chiuso nelle dinamiche di prossimità spaziale e temporale, ha dovuto fare i conti con il "non limite" e uscire fuori dal doppio recinto dello spazio di prossimità della vita corporea umana e da quello dello Stato caratterizzati da poche e lente novità perché da quel recinto era già uscita l'azione umana grazie alla tecnica dell'IA.

L'articolo esplora i ruoli paralleli delle norme giuridiche e degli algoritmi nella creazione di regolarità benefiche per la cognizione umana, proteggendola dall'imprevedibilità che è resa ancora più ampia (diremmo infinita) dalle condizioni relazionali dell'età globale e tecnologica.

Se la norma giuridica è per sua natura teleologica e il ragionamento causale ne è la base, la regola algoritmica potrebbe non esserlo. Il sistema di apprendimento automatico è infatti di natura non abduttiva e la conseguente apofenia dell'intelligenza artificiale, la capacità cioè di riconoscere schemi o connessioni tra informazioni che non hanno una significativa correlazione logica, se da un lato potrebbero rivelarsi una risorsa incrementale (si pensi all'invenzione dell'antibiotico Alicina) dall'altro, nelle scienze giuridiche l'immotivata visione di connessione sembra sfidare l'essenza stessa della regola giuridica fondata sulla legittimazione democratica e la sua limitazione costituzionale.

Si tenterà pertanto di chiarire, attraverso l'analisi delle varie declinazioni di opacità algoritmica, connesse alla natura non causale dei processi di apprendimento automatico, come il requisito della trasparenza, se isolatamente considerato, si riveli da solo insufficiente a determinare un'affidabilità dei sistemi di IA tali da renderli potenzialmente sostitutivo dell'*agency* umana. L'analisi dell'autonomia decisionale conferma l'impossibilità oggettiva di delegare la definizione del substrato etico-valoriale che fornisce contenuto agli assiomi su cui la decisione algoritmica è costruita. Come si cercherà di mettere in luce nel corso del terzo paragrafo, lo sforzo di educare l'algoritmo alla *fairness* sostanziale e non solo procedurale presuppone non soltanto la definizione di carattere squisitamente umano, dei principi etico-normativi a monte, formalizzati nel processo matematico, ma anche una "negoziazione" degli stessi quanto più consapevole e democratica. Data l'impossibilità epistemica di raggiungere un'utopica oggettività del procedimento matematico e l'assenza di regole non è e non potrà mai tradursi in agnosticismo valoriale, ma una delega in bianco ai loro programmatori. L'IA è creazione umana, creata a immagine e somiglianza dell'umano, e in quanto tale, nel bene e nel male, non potrà non incorporare valori, norme, pregiudizi e stereotipi dello spazio sociale specifico in cui processi cognitivi prendono forma, vestendone lo specifico *habitus* culturale. Se dunque la "verità algoritmica" non è che l'esito di una selezione tra più

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174

opzioni disponibili, più o meno condivisibili, allora il ruolo del regolatore è quello di rendere la negoziazione etico-normativa il più possibile democratica, agendo a vari livelli del processo algoritmico.

L'articolo sottolinea la necessità di approcciare alla regolazione dell'IA contemporaneamente con più strumenti: regolando i dati messi a disposizione, l'attività di elaborazione e l'output fornito. Se da un lato è ben valutabile l'attuale approccio europeo in materia (*AI-Act*) che si basa si un approccio basato sul rischio e costituiscono certamente una opportunità gli altri approcci che intervengono sul prodotto finale, sia per renderne trasparente l'origine sia per valutarne i contenuti con un accresciuto ruolo della capacità umana di valutazione degli stessi, l'articolo propone un orizzonte di intervento più ampio: intervenire sulla limitazione dei dati offribili all'IA. Parafrasando quando accaduto con il mercato che, agendo oltre gli Stati, ha costruito un ordine giuridico distinto da quello pubblico e in mano agli operatori privati che si impone all'uomo (contratti, prodotti e costi orami standardizzati), occorre che lo Stato decida il recinto di azione dell'IA visto che essa costruisce un ordine che si impone all'individuo e che esso non concorre a definire attraverso i canali tradizionali della legittimazione politica del costituzionalismo democratico.

 ISSN 2612-6672 | DOI 10.54103/2612-6672/22174