

*Repliche alla Lettera****Il Prometeo regolato: “l’uomo è antiquato”, o lo sono le istituzioni?*****di Francesco Cirillo**

Vorrei provare, in risposta solo ad alcuni dei molti interrogativi posti, a segnalare alcune criticità della regolamentazione dell’AI, che sono pur documentate in dottrina e però godono di minor fortuna nel dibattito (scientifico e pubblico). Nel contesto europeo, alcune tecnologie di trattamento dei dati, in particolar modo rivolte (anche) ai dati personali, che siano implementate nell’offerta di prodotti e servizi digitali online, incontrano oggi una trama regolatoria inconsueta, se non altro per la densità sempre crescente e per il ricorso a clausole e principi generali che spesso integrano la vaghezza con una scivolosa ambiguità (ho tentato di suggerire in questo forum – lett. 5/2023 – che lo stesso riferimento all’AI “in generale” mostri probabilmente più di un vulnus concettuale).

In ogni caso, tali applicazioni tecnologiche possono essere assoggettate contestualmente al GDPR, alla normativa ePrivacy, alle regole sulla cibersicurezza, al pacchetto sui servizi e sui mercati digitali (DSA e DMA), e al futuro regolamento sull’AI, senza che ciò le sottragga – secondo l’occasione – alle discipline degli Stati membri, nei casi in cui siano sussumibili nelle fattispecie regolate dai diversi settori dell’ordinamento. Se già la densità e l’estensione del quadro regolatorio possono istillare qualche dubbio sull’effettività, le perplessità non si risolvono se si osserva il contenuto e l’approccio dei singoli interventi regolatori.

Tra gli elementi più significativi e ricorrenti si registra l’insistenza sulla progettazione del software ispirata ai valori fondamentali, che andrebbero introiettati nel design delle piattaforme. Questa ricetta, secondo i suoi fautori, consentirà di sviluppare in Europa più che altrove una tecnologia incentrata sull’umano, in grado di bilanciare rischi e benefici, e di promuovere i diritti anziché minacciarli. Con l’effetto ulteriore (il cd. Brussels effect) di suscitare nella comunità globale il fascino delle più virtuose soluzioni europee.

E, tuttavia, l’ambizione presenta nelle finalità e nel metodo non poche oscurità. Si assiste, infatti, all’evocazione di una coesistenza irenica dei valori, che offusca il rilievo politico della decisione sul conflitto tra diritti e interessi (da cui dipende la prevalenza di un valore sull’altro). E quest’irenismo si avvale di una moltiplicazione ad libitum dei soggetti coinvolti nel bilanciamento tra diritti e interessi, sino a incorrere in una seconda classe di aporie. Nel sovrapporsi dei rinvii all’adeguatezza, all’idoneità, alla proporzionalità della tecnologia, il regolatore si affida, da ultimo, ai destinatari stessi delle regole, alla tecnologia o alla comunità scientifica. Non occorre soffermarsi sui pregi e sui difetti dei modelli autoregolatori e tecnoregolatori, su cui già tanto e bene è stato detto. In modo emblematico, mi limiterei a segnalare gli artt. 31 e 40 del Digital Services Act, i quali, facendo proprio

un approccio già del tutto consentaneo al GDPR, istituiscono e regolano la funzione “paranormativa” del «ricercatore abilitato»; una “versione 4.0” del filosofo-re, che richiama le speranze che François Géný affidava al potenziale nomopoietico della ricerca (p.e., in *Science et technique en droit privé positif*, 1924).

Ma, così facendo, il legislatore europeo, nel tentativo di affidare le decisioni che non sa o non può prendere a soggetti che sembrano più adeguati (per prossimità, competenze e rapidità decisionale), incontra le sabbie della fallacia naturalistica. Pretende, in altri termini, che la dimensione normativa, che gli appare sempre più necessitata e però ineffabile, sorga spontaneamente dal piano dei fatti. “Spera” che il dover essere nasca dall’essere. Così, impone allo sviluppatore del software di realizzare il bilanciamento dei diritti, adeguandosi a un corpus di regole che già si compone di migliaia di disposizioni e considerando (che però eludono le scelte sul conflitto, invocando un pantheon di valori, sempre ben bilanciati e proporzionati). Si tratta, forse, di un complesso di effetti distorsivi che gravitano attorno a un nucleo comune di questioni di fondo, proprio per quanto mi pare di aver (non appreso, ma quantomeno) intuito dagli itinerari di ricerca di chi ha introdotto la riflessione.

Molte di queste perplessità teoriche non tardano ad avere conferme empiriche. Al cospetto di questa trama normativa, le piattaforme (che dovrebbero esserne trasfigurate) continuano a operare in Europa non diversamente che nel resto del mondo, scoprendosi di volta in volta adeguate al quadro regolatorio o, al più, bisognose di qualche correttivo. Persino nei casi dove le scelte sono state compiute, e le regole sembrerebbero elevare un argine contro la tecnologia, le criticità non sembrano svanire. Si possono evocare le soglie sancite dal pacchetto sui servizi e sui mercati digitali, che delineano una disciplina di sfavore per un numero di destinatari predeterminabili (tutti o quasi extraeuropei), aprendo a un nuovo quadro di questioni.

A ben vedere, proprio la mescolanza tra aporie teoriche e inefficienze applicative pone in risalto le discrasie tra le finalità della regolamentazione e le sue conseguenze. La strategia digitale tenderebbe a promuovere, sulla base di interventi regolatori (e ora anche economici), un mercato digitale europeo sui generis. Eppure, questa strategia, oltre a manifestarsi soprattutto in blandi interventi protezionistici, non sembra dispiegare gli effetti sperati, né in termini di controllo degli effetti collaterali di una tecnologia che (sia assume o si presume) distorsiva e pervasiva, né in termini di promozione di un nuovo mercato in cui si affermino, finalmente, attori e tecnologie europee.

E, quanto alla proposta sull’AI, stando all’ultimo draft disponibile, sembrano confermate le perplessità: il richiamo irenico ai diritti (cons. 1-5, ecc.), la moltiplicazione dei decisori (si aggiungono, al quadro già articolato, l’autorità di notifica, gli organismi di valutazione delle conformità, ecc.), la fallacia naturalistica, la normazione per clausole vaghe, se non ambigue. Su quest’ultimo punto, si osservi che la disciplina, che forse vorrebbe rivolgersi solo ad alcune applicazioni, si interessa piuttosto di classi di operazioni statistico-computazionali, del tutto neutre se considerate in generale. Non è chiaro, infatti, in che senso si assuma il rischio delle stime “bayesiane” (all. 1, lett. c), se non addirittura degli «approcci basati sulla logica e approcci basati sulla conoscenza» e degli «approcci statistici» (lett. b e c). Immaginiamo, per assurdo, che il legislatore

mostri uno sfavore verso le equazioni di terzo grado, per via della loro intrinseca attitudine ad apparire oscure all'utente medio.

Ma, soprattutto, non è chiaro in che modo una regolazione così intensa possa impedire all'utente di avvalersi comunque, con le opportune soluzioni tecniche, di tecnologie extraeuropee, che già oggi presentano un netto vantaggio competitivo e godrebbero sempre più di spazi in cui svilupparsi senza gravami. Si pensi, in tal senso, alla possibilità di aggirare i divieti locali per via del ricorso a strumenti facilmente accessibili, quali, per esempio, le VPN (non sia un'istigazione il richiamo a quanto avvenuto in conseguenza del blocco temporaneo di ChatGTP).

Sembrano in parte svalutate, almeno nel dibattito pubblico, proprio quelle «straordinarie opportunità di progresso»: le potenzialità che i «sistemi di AI» possono dispiegare nell'efficientamento delle infrastrutture, nell'organizzazione degli spazi urbani, nel contesto dell'energia, nella medicina (preventiva, diagnostica, interventistica, ecc.), ma anche solo per quel che concerne – prima ancora dello sviluppo di processi decisionali automatizzati – il rafforzamento della conoscenza di fenomeni complessi che ambiscono a uno studio statistico-computazionale (le epidemie, i flussi migratori, i sistemi giudiziari, i sistemi formativi, l'attività della PA, il mercato del lavoro, ecc.).

Si obietterà che questa prospettiva sottovaluti i rischi, le distorsioni, le compressioni di diritti, i bias che la tecnologia alimenta o produce. Nondimeno, non è parso utile aggiungere una voce al coro delle denunce dei rischi – che pure risultano spesso fondate e condivisibili –, quanto piuttosto evidenziare che uno stravolgimento della postura europea sul tema della ricerca, sui rapporti tra Stato e mercato, tra innovazione e regolamentazione, potrebbe meritare la stessa prudenza di quanta se ne dimostri per le tecnologie emergenti. Sempre più spesso, infatti, assistiamo a una torsione delle istituzioni nei confronti della ricerca e del mercato, con forti elementi di dirigismo, di pianificazione o di accentramento (per tutti, si pensi al PNRR o alle politiche ambientali).

In questo scenario, forse, andrebbe almeno registrato, insieme con gli altri, un rischio ulteriore e non secondario: che l'Europa dei diritti perda (anche, e da ultimo) la sfida dell'AI, e sia destinata per sua scelta a essere soprattutto un museo delle virtù e degli orrori del passato, uno spazio immobile, sopraffatto da vacue speranze e da paure insormontabili.

Prometeo non era un dinosauro: diritto costituzionale e A.I.**di Gian Luca Conti**

1 – Il punto di partenza di Luciani è una idea “catastrofica” del futuro. Lo sviluppo tecnologico, e, in particolare, lo sviluppo della intelligenza artificiale, si potrebbe sottrarre al controllo etico dell’uomo e potrebbe distruggere la stessa umanità, almeno come la conosciamo.

Di qui, il richiamo al concetto di “dislivello prometeico” (G. Anders, *Die Antiquiertheit des Menschen. Über die Seele im Zeitalter der zweiten industriellen Revolution*, München, 1956, trad. it. *L’uomo è antiquato*, vol. I, Torino, Bollati Boringhieri, 2003², part. 50).

L’approccio catastrofista, a ben vedere, si trova agganciato a un’antropologia del moderno in cui il principio liberale ha determinato il sorgere di un individualismo illimitato: l’uomo che crede solo in se stesso ma si chiude in comunità che si autoalimentano esclusivamente della propria cultura identitaria di talché la paura diventa il vettore di senso del legame sociale (in questi termini: E. Pulcini, *La cura del mondo. Paura e responsabilità nell’età globale*, Torino - Bollati Boringhieri, 2009).

È un circolo vizioso in cui l’*homo faber* di Anders è governato dalle proprie emozioni e la più forte di queste (la paura) lo conduce a produrre il proprio olocausto.

Questo tipo di catastrofismo si pone come una critica radicale alla cultura liberale una delle cui massime è vivere pericolosamente: «Non c’è liberalismo senza cultura del pericolo» (M. Foucault, *Naissance de la biopolitique: Cours au collège de France (1978-1979)*, Paris, Seuil, 2003, trad. it. *Nascita della biopolitica. Corso al Collège de France (1978-1979)*, Milano, Feltrinelli, 2005, part. 68-69, su cui J. Donzelot, *Michel Foucault et l’intelligence du libéralisme*, in *Esprit*, 68, 2005, n. 10, pp. 56-74).

Il catastrofismo rivela la crisi dell’idea di Costituzione come strumento di apertura al possibile, secondo le intuizioni di Bataille (G. Bataille, *Recensione a K. Jaspers, Nietzsche. Einführung in das Verständnis seines Philosophierens*, Berlin 1936, in Id., *La congiura sacra*, Torino - Bollati Boringhieri, 1997, pp. 51-53, p. 52: «nella ricchezza del possibile, senza principio razionale, il contrario e il contraddittorio possono essere tentati, tentativo che obbedisce al solo principio della salvezza e della valorizzazione della condizione umana»).

La visione catastrofica di Anders è affascinante e scoraggiante: che spazio per il diritto costituzionale in un mondo in cui l’uomo rinnega la propria umanità perché ha imparato a farne a meno grazie all’aiuto di macchine sempre più complesse nei confronti delle quali soffre un complesso di inferiorità?

In Unione Sovietica si guardava con interesse e polemico senso di superiorità agli sviluppi delle scienze cibernetiche nei paesi del Patto Atlantico – l’intelligenza artificiale, come internet, nasce, nel

clima della guerra fredda, dalle intuizioni matematiche di Turing e Oettinger ma anche dai lavori pionieristici sulla cibernetica di Licklider (J.C.R. Licklider, *Man-computer symbiosis*, *IRE Transactions on Human Factors in Electronics*, 1960, HFE-1, pp. 4–11, facilmente reperibile in rete) –: «L'intelligenza tecnica occidentale è impigliata nel problema “uomo-macchina” perché non sa come formularlo adeguatamente; cioè come problema sociale, come problema del rapporto tra uomo e uomo, mediato dal corpo materiale della civiltà, compresa la moderna tecnologia meccanica della produzione» (Evald Ilyenkov, Aleksandr Arsen'ev and Vassily Davidov, *Mashina i chelovek: kibernetika i filosofiya* (Machine and human: cybernetics and philosophy), in Fedor Konstantinov (ed.), *Leninskaya Teoriya Otrazheniya i Sovremennaya Nauka*, Moscow: Nauka, 1966, pp. 265–83, cit. in O. Kirtchik, *The Soviet scientific programme on AI: if a machine cannot 'think', can it 'control'?*, in *Temi BHJS*, 2023, 8, 111-125).

Il problema è come cambia il rapporto fra gli individui e fra l'uomo e l'ambiente che lo circonda quando tale rapporto è mediato da una macchina capace di risolvere problemi complessi.

Il conducente del furgone che ci ha consegnato gli ultimi libri che abbiamo acquistato online prova risentimento verso l'intelligenza artificiale, verso l'organizzazione sociale che lo costringe ad accettare quel lavoro o verso chi, con spregiudicata intelligenza tutta umana, ha saputo sfruttare la sua situazione di bisogno avvalendosi della tecnologia?

La risposta razionale è facile e spinge a considerare l'intelligenza artificiale come un problema di diritto costituzionale la cui peculiarità è la capacità di modificare in maniera rilevante i termini del contratto sociale.

2 – Con notevole approssimazione, è possibile definire l'intelligenza artificiale come un sistema capace di assorbire informazioni e di utilizzarle con una capacità di risolvere problemi il più possibile vicina a quella umana (in termini più ampi, ISO/IEC 42001:2023 Information technology - Artificial intelligence Management System (AIMS) definisce l'intelligenza artificiale come la capacità di un sistema di mostrare capacità umane quali il ragionamento, l'apprendimento, la pianificazione e la creatività).

Per essere precisi, si dovrebbe tracciare una genealogia del concetto di intelligenza artificiale (genealogia nel senso di M. Foucault, *Hommage à Jean Hyppolite* [*Nietzsche, la généalogie, l'histoire*], in *Hommage à Jean Hyppolite*, Paris – Presses Universitaires de France, 1971, pp. 145-172], trad. it., in M. Foucault, *Microfisica del potere. Interventi politici*, Torino, Einaudi, 1977, pag. 29 e ss.). Non sempre si fa quello che si dovrebbe e qui conviene concentrarsi sul deep learning.

Il deep learning lavora tramite reti neurali in cui ciascun livello stabilisce che cosa può passare al livello successivo, ovviamente più sono i livelli e più il sistema è accurato.

Il test CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart), quel test che chiede di distinguere gli alberi dai semafori quando si accede a un sito internet, ha consentito di raccogliere una grandissima mole di dati da organizzare in reti neurali.

Queste reti neurali si sono evolute in due direzioni. La prima sfrutta un diagramma Bayesiano (ovvero un modello grafico probabilistico che rappresenta un insieme di variabili stocastiche con le loro dipendenze condizionali attraverso l'uso di un grafo aciclico diretto): se voglio capire perché la

ma una macchina non si accende mi devo porre una serie di problemi (variabili stocastiche), ciascuno dei quali impone una dipendenza condizionale che a sua volta può essere considerata come una variabile stocastica che determina ulteriori dipendenze condizionali. Questo diagramma consente di collegare le parole fra di loro seguendo il significato che le rende parti di un unico discorso e per ottenere questa coerenza classifica le parole come elementi di un insieme statistico di cui il significante è la variabile mentre il significato è la dipendenza, cosa che viene fatta utilizzando lo stesso modello (le catene di Markov) che si usa per predire il tempo o indovinare i numeri della roulette, un modello in cui una serie di variabili stocastiche non dipende necessariamente dalle dipendenze condizionali che le hanno generate e, perciò, in cui è possibile arrivare allo stesso punto (alle stesse conclusioni) indipendentemente dal punto di partenza.

Niente di semplice, insomma. Ma comunque un sistema in cui il nodo è combinare le variabili stocastiche alle dipendenze condizionali e che può essere vincolato considerando talune variabili come invarianti.

La seconda considera ogni parola come un insieme di significati che può essere misurato esattamente come se fosse una forza fisica, sicché il risultato di una combinazione di parole è un vettore che risulta da più forze semantiche.

Questa intelligenza artificiale funziona molto bene con i dati biometrici perché se guardo un viso e lo misuro, millimetro per millimetro posso ottenere il vettore che definisce solo quel viso. Il sistema di riconoscimento facciale di un telefono portatile funziona così.

L'intelligenza artificiale merita di uscire dal suo nascondiglio fantascientifico, di essere presa sul serio e compresa nel suo concreto funzionamento in modo da poter essere regolata.

Il modello Bayesiano, ovviamente, può essere disciplinato trasformando alcune variabili in valori costanti.

Il modello vettoriale, invece, può essere disciplinato a partire dai dati utilizzati per alimentarlo.

Ma il vero dislivello prometeico non riguarda il funzionamento della intelligenza artificiale, bensì il senso che acquista la macchina nelle relazioni fra gli uomini.

L'intelligenza artificiale deve essere regolata per il modo in cui può essere utilizzata: il programma utilizzato per la gestione dei conducenti di una piattaforma di commercio elettronico può essere impostato uno strumento di oppressione o un cronotachigrafo intelligente.

L'intelligenza artificiale più affascinante, però, è quella che può programmare se stessa, sia correggendo i difetti della sua programmazione, sia sviluppando la propria programmazione per affrontare i problemi che non erano stati immaginati quando era stata messa in opera ed arrivare fino a definire il proprio codice sorgente o addirittura il linguaggio di programmazione con cui opera. Fa paura perché non è più una invenzione umana.

Questa macchina che pensa (quasi) come un uomo può essere cattiva come un uomo?

Ma se fosse così, se avessimo effettivamente compreso il modo di ragionare delle macchine intelligenti per come si sta evolvendo e sviluppando, allora il monito con cui Luciani conclude la sua riflessione dovrebbe essere ripensato: non dobbiamo sforzarci di educare i nostri studenti a studiare con serietà, evitando scorciatoie, ma dobbiamo iniziare a pensare come educare le macchine "pensanti".

Il vero salto della intelligenza artificiale si è compiuto quando si è sostituito un pensiero debole in cui le variabili, i nodi, erano predeterminati a un pensiero forte in cui la stessa macchina è in grado di costruire le proprie variabili e i propri nodi.

Ecco questo non dovrebbe essere possibile, non del tutto. Ci sono dei nodi che devono essere predeterminati.

Ma per arrivare a questo si deve seguire la stessa strada della IRTF (Internet Research Task Force) che da tempo cerca di integrare i diritti umani nei protocolli di rete (vedilo a <https://hrpc.io/>), il che è relativamente semplice quando si parla di privacy perché è sufficiente disciplinare il linguaggio di programmazione delle pagine web in modo che mantenga privati i dati relativi alla navigazione, ma molto più complesso quando si parla di libertà di manifestazione del pensiero.

Non è la strada che segue l'Unione Europea o l'executive order di Biden citati da Luciani, che preferiscono guardare alle applicazioni, con una impostazione assai più prudente e tradizionale, perché basata su di una analisi del rischio che opera solo ex post e che, nello stesso tempo, non approfondisce i complessi temi che l'intelligenza artificiale pone sul piano della responsabilità aquiliana (cfr. C. Wendehorst, *Liability for Artificial Intelligence: The Need to Address Both Safety Risks and Fundamental Rights Risks*, in S. Voeneky, P. Kellmeyer, O. Mueller, W. Burgar, eds. *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives*, Cambridge Law Handbooks. Cambridge University Press; 2022, 187-209, anche per gli accenni alla disciplina in itinere a livello europeo).

3 – Sulla base di queste premesse, è possibile tentare qualche profilo di risposta ai quesiti posti da Luciani.

Il primo di questi riguarda la possibilità di considerare l'intelligenza artificiale come un fattore di rischio.

Il rischio è lo strumento con cui l'uomo cerca di stare meglio o meno peggio (A. Smith, *Theory of Moral Sentiments* (1759); trad. it. *Teoria dei sentimenti morali*, Roma, Istituto dell'Enciclopedia Italiana, 1991).

Il valore della libertà scientifica per come è espresso nell'art. 33, Cost. può ricollegarsi alla matrice liberale della libertà di ricerca? La Costituzione non fugge dal rischio, quando afferma la libertà della ricerca, è consapevole del rischio che la ricerca produce: il Costituente sapeva perfettamente che la scienza era pericolosa, era molto più vicino di noi agli sviluppi letali delle ricerche in campo nucleare.

Nello stesso tempo, la proprietà, anche la proprietà intellettuale, non solo quella fondiaria o dei mezzi di produzione, è costituzionalmente intrisa di funzione sociale e questo vale anche per il codice dell'intelligenza artificiale. Lo stesso vale per l'iniziativa economica, anche nel campo delle nuove tecnologie, deve essere indirizzata verso l'utilità sociale e, adesso, lo sviluppo sostenibile.

In altre parole, è chiara, a mio avviso, nel testo costituzionale l'opzione per una libertà della ricerca che prelude a una società del rischio e l'idea che questo debba essere attentamente orientato da un'azione pubblica possibile per effetto del complesso coordinarsi degli artt. 41, 42 e (perché no?) 43, Cost.

Ma Luciani pone anche il problema radicale della possibilità logica di regolare la ricerca scientifica.

In realtà, il diritto costituzionale è consapevole che senza una società del rischio sarebbe molto difficile modificare il senso dei rapporti sociali nella chiave dell'eguaglianza, e, nello stesso tempo, consente di distinguere il dominio, in senso matematico, della ricerca scientifica dal dominio della tecnologia che cade negli artt. 41 – 43, Cost.

La ricerca scientifica, come ogni altro valore costituzionale, non può essere considerata un valore assoluto, non lo sono neppure l'ambiente o la salute, sicché ben può accadere che sia regolata per tenere conto di altri valori e questo mi pare valga più con riferimento alla scienza applicata (la produzione di embrioni, cui accenna Luciani) che non alla scienza pura (la ricerca delle condizioni alle quali è possibile riprodurre sinteticamente un essere umano).

Luciani chiede che cosa il diritto dovrebbe promuovere con riferimento alla scienza e in particolare alla ricerca nel campo della intelligenza artificiale.

Si è già risposto cercando di spiegare che cosa è l'intelligenza artificiale e individuando nella sua mancanza di "educazione" il problema. Lo Stato deve incoraggiare una ricerca capace di educare l'intelligenza artificiale generativa. Ma questo è giusto? Non diventa una sorta di censura? Se la cifra del contemporaneo è la "governamentalità", ancora il Foucault della biopolitica, l'intelligenza artificiale asservita allo Stato non potrebbe condurre a un assolutismo catastrofico di valori che sarebbero condivisibili solo se "parlamentati" con metodo democratico?

Se il diritto sceglie di regolare esclusivamente le applicazioni, questa regolazione, seguendo lo schema della disciplina eurounitaria o di quella emanata negli Stati Uniti, non può che farsi carico di una adeguata analisi del rischio connesso a ciascuna applicazione e muoversi di conseguenza, se, invece, il diritto si prende carico della intelligenza artificiale considerandola come uno strumento che può servire per promuovere i valori costituzionali, allora, lo sforzo diventa incorporare i valori costituzionali nel codice dell'intelligenza artificiale, nel suo patrimonio genetico, facendone delle invarianti.

E questa, ad avviso di queste pagine, dovrebbe essere la sfida che il diritto costituzionale dovrebbe porsi dinanzi all'intelligenza artificiale: se gli uomini possono essere cattivi, le macchine pensate dagli uomini possono essere programmate in modo da essere, se non buone, meno cattive?

4 – L'intelligenza artificiale spaventa perché il suo sviluppo può sfuggire al controllo umano.

L'uomo – l'uomo comune, quello fatto come ciascuno di noi – guarda l'intelligenza artificiale con il muto spavento che spingeva l'antico greco a rivolgersi agli indovini per placare gli dei, perché sa di essere inferiore, sa di soffrire del dislivello prometeico sofferto dagli ultimi dinosauri prima della loro estinzione.

Ma se c'è una cosa che la modernità occidentale ha compreso è che il rischio deve essere considerato inevitabile e che la sua inevitabilità produce paura, quella paura che è il vettore di senso dei legami sociali a partire da Hobbes.

La Costituzione è una risposta a questa paura perché costituisce un vettore di senso dei legami sociali capace di spiegare e rendere ragionevoli le diseguaglianze, è lo strumento con cui il potere costituente indirizza la società verso il possibile.

Se Prometeo fosse stato un dinosauro, gli uomini non avrebbero mai avuto il fuoco. La sua paura di estinguersi avrebbe vinto su tutto, ma se Prometeo fosse stato un costituzionalista prima di consegnare il fuoco avrebbe pensato a come il fuoco avrebbe potuto essere utilizzato per ridurre le diseguaglianze invece che per vincere le guerre.

Il problema dell'intelligenza artificiale non è più un problema che si può risolvere sul piano della cibernetica, è un problema antropologico, perché è antropologica la risposta costruttiva ai dilemmi posti dalla società del rischio (E. Pulcini, *Soggetti emozionali. Per avere cura del mondo*, in *Iride*, 2021, 267) e, forse, è proprio la dimensione antropologica (e curativa) della Costituzione che dovrebbe essere indagata per trovare in maniera più approfondita e costruttiva le risposte ai quesiti che questa lettera ci ha posto.

Senza cedere alle tentazioni del catastrofismo: se i dinosauri avessero potuto decidere, non avrebbero mai acconsentito all'avvento dei mammiferi e l'essenza del mito di Prometeo è che l'uomo non è un dinosauro.

Il valore della trasparenza nei sistemi sociali algoritmici: questioni di ordine linguistico e costituzionale

di Maria Romana Allegri

Massimo Luciani, nel secondo paragrafo del suo contributo introduttivo al tema “libertà di ricerca e intelligenza artificiale”, ha prospettato alcuni interrogativi che vertono intorno al rapporto fra diritto e scienza: è possibile e opportuno che il diritto regoli la ricerca scientifica o almeno le sue applicazioni, con particolare riferimento ai sistemi di intelligenza artificiale? A tale proposito, Francesco Cirillo, nella sua replica, ha evidenziato alcune criticità: le regole prodotte a livello di Unione europea in riferimento alle tecnologie digitali (il GDPR, i recenti regolamenti sui servizi digitali e sui mercati digitali, regolamento *in fieri* sull’intelligenza artificiale) ricorrono spesso a concetti vaghi e ambigui, producendo effetti distorsivi (fra cui quello dell’abnorme moltiplicazione dei decisori) e risultando in buona parte inefficaci. Da queste considerazioni possono derivare alcuni spunti di riflessione cui mi sembra opportuno accennare in questo spazio di dibattito relativamente informale, pur consapevole della necessità di ulteriore studio e approfondimento per rendere il ragionamento un po’ meno semplicistico. Mi vorrei soffermare, in particolare, sulla questione della trasparenza algoritmica nella più recente normativa e giurisprudenza, che si scontra con l’impossibilità di tradurre i processi algoritmici in linguaggio comune.

Come evidenziato in varie occasioni da Andrea Simoncini (anche in *Rivista AIC*) oggi i sistemi di intelligenza artificiale – a prescindere dalle varie e non sempre univoche definizioni di AI fornite in vari momenti e a vari livelli – sono diventati una parte essenziale della vita quotidiana, tanto da poter parlare di *social machine* oppure di “sistema sociale algoritmico” per indicare quei sistemi che nascono dalla strutturale interazione di elementi umani e tecnici. Viviamo dunque in un ecosistema basato sull’interazione uomo-macchina e non a caso già da qualche anno il Consiglio d’Europa e il Parlamento europeo hanno sottolineato, in alcune risoluzioni, i rischi in termini di manipolazione della coscienza e della volontà connessi allo sviluppo dei sistemi tecnologici di interfaccia *brain-computer*. Peraltro, il concetto di rischio è il cardine attorno al quale ruota il regolamento europeo sull’intelligenza artificiale, ormai in dirittura d’arrivo.

Ma, al di là degli specifici rischi connessi alle singole applicazioni delle tecnologie di AI, qual è il rischio di fondo sotteso all’intero sistema? Mi pare che ciò consista nella perdita di centralità dell’elemento antropologico nell’attuale ecosistema, posto che – sempre citando Simoncini – «siamo dinanzi ad una delega (totale o parziale) di funzioni cognitive ritenute sinora proprie ed esclusive degli esseri umani, a sistemi tecnologici»: la tecnologia non è più soltanto un mezzo utilizzato per eseguire in modo più efficiente decisioni prodotte dall’uomo, ma è il soggetto agente che prende autonomamente decisioni in grado di incidere nella sfera giuridica delle persone umane, attraverso

processi di elaborazione di dati. Gli algoritmi definiscono e limitano gli orizzonti epistemici dell'individuo, poiché plasmano e determinano le modalità con cui esso entra in contatto con il mondo, le sue opportunità di conoscenza e le sue alternative di scelta, fino ad interferire con la sua libertà morale e con la stessa costruzione identitaria del sé. Ciò mette in crisi il fondamento stesso del diritto costituzionale, basato sulla centralità dell'individuo e sulla garanzia dei suoi diritti, se non del diritto *tout court* inteso come tecnica per il controllo del comportamento dell'uomo, che si realizza attraverso il ricorso ad enunciati linguistici.

Ora, è vero che qualsiasi apparato tecnologico deriva originariamente dall'atto creativo umano e che dunque all'origine dei sistemi di AI vi è l'attività umana di progettazione degli algoritmi alla base del loro funzionamento. Tuttavia, le dinamiche inferenziali e statistico-probabilistiche attraverso cui il sistema tecnologico "impara" dai dati sono tali per cui la "logica" alla base delle decisioni assunte dalla macchina non è comprensibile per l'uomo, nemmeno per i programmatori degli algoritmi stessi: non è spiegabile prima che la decisione algoritmica venga assunta né ricostruibile a posteriori ripercorrendo a ritroso la catena dei processi di elaborazione. È così, di fatto, perché le decisioni algoritmiche, a differenza di quelle che ispirano l'agire umano, sono basate sul calcolo delle probabilità e non su processi logico-deduttivi in cui il giudizio è conseguenza della valutazione dei presupposti in termini non solo utilitaristici, ma anche etici.

Tutto ciò considerato, ci si chiede come possa realizzarsi in concreto il principio della trasparenza algoritmica, che è un *mantra* ricorrente nella normativa dell'Unione europea in ambito digitale. Si pensi, ad esempio, al richiamo a «informazioni significative sulla logica utilizzata» contenuto nell'art. 15, par. 1, lett. h del GDPR, oppure all'art. 27 del regolamento europeo sui servizi digitali, che postula la trasparenza dei sistemi di raccomandazione da realizzarsi «in un linguaggio chiaro e intellegibile», oppure ancora ai frequenti riferimenti alla trasparenza dei sistemi di AI nel regolamento europeo di prossima approvazione, che dovrebbero far sì che l'utente sia in grado di comprendere e utilizzare adeguatamente tali sistemi avendo una conoscenza generale del loro funzionamento. La trasparenza, di fatto, è irrealizzabile, non solo per l'impossibilità di spiegare le motivazioni alla base delle decisioni algoritmiche, ma anche perché, più semplicemente, tradurre il *software* nel linguaggio comune comporta inevitabilmente un alto grado di imprecisione e di approssimazione. Dunque, quale livello di comprensibilità dell'algoritmo può essere considerato adeguato a integrare il requisito della trasparenza?

Infatti, nemmeno l'ostensione del codice sorgente dell'algoritmo, cioè del *software* elaborato con il linguaggio di programmazione – come ad esempio ha prescritto il Tar Lazio, sez. III bis, nella sentenza n. 3769 del 2017, relativa all'algoritmo utilizzato per formare le assegnazioni alle sedi di servizio degli insegnanti – risolverebbe di per sé il problema della trasparenza algoritmica, non solo perché difficilmente comprensibile per i non addetti a lavori, ma anche perché non necessariamente riuscirebbe a spiegare le ragioni alla base di una decisione assunta da algoritmi di *machine learning*, progettati per "imparare" dai dati ed evolversi automaticamente di conseguenza.

Secondo il Consiglio di Stato (sez. IV, sentenza n. 2270 dell'8 aprile 2019) il principio della trasparenza algoritmica implica la «piena conoscibilità di una regola espressa in un linguaggio differente da quello giuridico». Quindi occorre che «la "formula tecnica", che di fatto rappresenta

l'algoritmo, sia corredata da spiegazioni che la traducano nella "regola giuridica" ad essa sottesa e che la rendano leggibile e comprensibile, sia per i cittadini che per il giudice». In particolare, questa «conoscibilità dell'algoritmo deve essere garantita in tutti gli aspetti: dai suoi autori al procedimento usato per la sua elaborazione, al meccanismo di decisione, comprensivo delle priorità assegnate nella procedura valutativa e decisionale e dei dati selezionati come rilevanti». Solo così, infatti, il giudice potrà «sindacare la stessa logicità e ragionevolezza della decisione amministrativa robotizzata, ovvero della "regola" che governa l'algoritmo». Il problema di fondo di questa decisione è che essa esplicitamente muove dall'erroneo presupposto che «la regola tecnica che governa ciascun algoritmo resta pur sempre una regola amministrativa generale, costruita dall'uomo e non dalla macchina, per essere poi (solo) applicata da quest'ultima»; tuttavia oggi i sistemi di intelligenza artificiale non si limitano ad applicare regole precostituite, ma riescono a produrre regole nuove sulla base dell'esperienza acquisita tramite di processi di elaborazione dei dati.

Proprio a tale proposito, successivamente il Consiglio di Stato (sentenza n. 7891 del 25 novembre 2021) ha distinto fra la nozione di algoritmo – «una sequenza finita di istruzioni, ben definite e non ambigue, così da poter essere eseguite meccanicamente e tali da produrre un determinato risultato» – e quella di intelligenza artificiale, in cui «l'algoritmo contempla meccanismi di *machine learning* e crea un sistema che non si limita solo ad applicare le regole *software* e i parametri preimpostati (come fa invece l'algoritmo "tradizionale") ma, al contrario, elabora costantemente nuovi criteri di inferenza tra dati e assume decisioni efficienti sulla base di tali elaborazioni, secondo un processo di apprendimento automatico».

Più o meno in linea con il Consiglio di Stato, la Corte di Cassazione (I sez. civ., ordinanza n. 28358 del 10 ottobre 2023) ha ritenuto soddisfatti i requisiti di trasparenza e conoscibilità dell'algoritmo (nella fattispecie, un algoritmo di *rating* reputazionale) nel momento in cui si pone l'utente in condizione di comprendere lo «schema esecutivo dell'algoritmo», cioè la sequenza dei passaggi necessari a raggiungere un risultato, partendo da determinati dati: tali passaggi devono essere «elementari, univoci, di numero finito, operabili in un tempo finito e con risultato unico» e devono essere spiegati all'utente «in modo non ambiguo e in maniera dettagliata» utilizzando i termini della lingua comune e non certo il linguaggio matematico, che l'utente non è tenuto a comprendere. Ma il problema resta: è davvero possibile descrivere senza ambiguità e con un sufficiente livello di dettaglio dei procedimenti matematici in linguaggio comune? E quali indicatori consentirebbero di valutare, in caso di controversie, se tale spiegazione sia adeguata? Ad esempio, proprio nel caso di specie il giudice di primo grado (tribunale di Roma) aveva ritenuto che il livello di trasparenza algoritmica non fosse soddisfatto poiché la sua descrizione si limitava al l'incidenza dei dati presi in considerazione, senza spiegare le modalità con cui si giungeva al risultato finale (cioè, il "peso specifico" dei singoli fattori considerati e i meccanismi di interazione fra loro).

Tale obiezione, per quanto non condivisa dalla Cassazione, è indicativa a mio parere di un problema insolubile: la spiegazione "a parole" dell'algoritmo produce inevitabilmente un certo grado di oscurità e ambiguità. Forse un rimedio, per quanto parziale, potrebbe consistere nella definizione, per via normativa, di indicatori e parametri il più possibile condivisi, in base ai quali misurare il livello di trasparenza algoritmica. Forse qualche soluzione operativa potrebbe provenire, in un

prossimo futuro, dal Centro europeo per la trasparenza algoritmica (ECAT), inaugurato il 18 aprile 2023, che riunisce ricercatori ed esperti nell'ambito dell'informatica, dell'analisi dei dati, dell'intelligenza artificiale e delle scienze giuridiche e sociali, al fine di analizzare il funzionamento dei *software* delle grandi piattaforme e motori di ricerca.

Comunque, nonostante tutte queste difficoltà definitorie e applicative, poiché è un dato di fatto che tutti noi siamo parte di un ecosistema “tecno-antropologico”, se non addirittura “antropo-tecnocratico”, nel quale siamo contemporaneamente progettisti-produttori di sistemi tecnologici e destinatari passivi e spesso inconsapevoli della tecnoregolazione, risolvere la questione della trasparenza e della comprensibilità dei *software* algoritmici e di AI è una priorità ineludibile, se davvero si vuole mantenere la persona umana al centro dei processi decisionali.

Ciò potrebbe, addirittura, condurre a un'interpretazione evolutiva della Carta costituzionale, a cominciare dall'art. 2 Cost.: i sistemi di interazione uomo-macchina ragionevolmente potrebbero essere considerati oggi come formazioni sociali in cui l'uomo svolge la sua personalità, in cui dunque devono essere garantiti i diritti fondamentali dell'individuo e deve essere richiesto l'adempimento dei doveri inderogabili di solidarietà economica, politica e sociale, gravanti anche sui gestori delle infrastrutture tecnologiche su cui tali sistemi si basano. Inoltre, in riferimento all'art. 3 Cost., la difficile intellegibilità dei linguaggi algoritmici da parte del cittadino comune rischia sempre più frequentemente di caratterizzarsi come un ostacolo che limita di fatto la libertà e l'eguaglianza dei cittadini, impedendo il pieno sviluppo della persona umana e l'effettiva partecipazione di tutti i lavoratori all'organizzazione politica, economica e sociale del Paese: infatti, essa può generare disparità, incompatibili con il principio di uguaglianza sostanziale, fra chi possiede gli strumenti cognitivi necessari a comprendere i linguaggi algoritmici e chi non ne dispone; può far sì che molte persone vengano escluse non solo dalla comprensione dei processi decisionali ma anche dall'accessibilità ai circuiti di *decision making*, producendo un *vulnus* sotto l'aspetto della loro piena partecipazione attiva alla vita democratica; può esporre le persone a lesioni della loro dignità personale o della loro *privacy*, senza che esse ne siano consapevoli e siano poste in condizioni di avvalersi efficacemente degli strumenti di tutela. Anche l'art. 9 Cost. nella sua più recente formulazione, potrebbe essere esteso agli ecosistemi basati sull'interazione uomo-macchina: una lettura, questa, in qualche modo favorita dal richiamo alla ricerca scientifica e tecnica contenuto nel primo comma e a quello del secondo comma riferito alle future generazioni, che ancora più di quelle attuali saranno plasmate dalla tecnologia. Poiché la tutela di cui all'art. 9 si riferisce agli ecosistemi in sé – quindi a tutte le loro componenti e non a quella umana in particolare – la norma potrebbe prestarsi non solo alla protezione dei diritti individuali all'interno dell'ecosistema digitale, ma anche alla protezione dello sviluppo tecnologico, essenziale alla sussistenza dell'ecosistema stesso. Infine, l'art. 24 Cost., che assicura il diritto di difesa in ogni stato e grado dei procedimenti giudiziari, potrebbe essere inteso come intrinsecamente collegato alla conoscibilità e comprensibilità dei processi algoritmici, senza cui sarebbe impossibile ricorrere efficacemente avverso qualsiasi decisione derivante da essi.

Certamente queste ultime considerazioni si prestano facilmente ad essere tacciate di “fantacostituzionalismo”. Ma se la nostra Costituzione non si aprirà in qualche modo alle sfide della

contemporaneità – come certamente ha fatto nel momento in cui è stata originariamente concepita – e se il diritto non si sforzerà di individuare nuovi percorsi, anche a rischio di finire in qualche campo minato, le generazioni future rischiano di vedere entrambi, fra non moltissimo tempo, relegati fra i reperti archeologici.