fuori collana

L’IA, quando il mezzo diventa soggetto

Federico Losurdo
19 Novembre 2024



Di fronte agli straordinari progressi dell’IA, in particolare di quella basata sul c.d. apprendimento automatico (*machine learning*), è necessario adottare un approccio equilibrato e senza preconcetti. Da un lato, occorre evitare visioni *apocalittiche* su un futuro imminente in cui le macchine ci domineranno (come esemplificato dall’ultimo, peraltro, egregio lavoro di Yuval Noah Harari, 2024). Dall’altro è altrettanto importante diffidare di visioni troppo *ireniche* che minimizzano i rischi sottesi ad uno sviluppo senza limiti etici dell’IA.

Il salto qualitativo dell’I.A.

La scrittura era ritenuta da Platone come una tecnologia “aliena”, allo stesso modo in cui oggi molte persone percepiscono l’IA». Tuttavia, l’invenzione del documento scritto ha consentito di creare un ordine sociale e dare certezza giuridica ai rapporti tra privati (per esempio con riguardo ai confini tra le rispettive proprietà). Allo stesso modo la *stampa* infranse il monopolio della Chiesa sulla verità, creando le premesse per la rivoluzione scientifica, ma fu anche veicolo delle prime campagne di “fake news” ante-litteram (come nel caso della “caccia alle streghe”, Harari, 2024). A loro volta, la radio e la televisione si

sono rivelate due formidabili tecnologie dell'informazione che hanno favorito l'evoluzione in senso democratico di molti ordinamenti, pur accompagnandosi, come contrappasso, alle derive totalitarie di altri Stati.

L'ultima tecnologia dell'informazione in ordine del tempo, il *personal computer*, ha fin dall'inizio evidenziato una straordinaria potenzialità che la differenzia da tutte le precedenti : «l'idoneità ad *assumere decisioni* e generare *idee* in relativa "autonomia» (Harari, 2024). Già negli Anni Cinquanta Alan Turing predisse un futuro in cui sarebbe stato difficile distinguere l'uomo dalla macchina, un futuro mirabilmente rappresentato nel capolavoro letterario di Philip K. Dick.

Prima di soffermarsi sulle implicazioni giuridiche e costituzionali dell'IA è utile definirne la "natura". Per semplificare, i "vecchi" algoritmi si basavano su un approccio *logico-deduttivo*. Il programmatore predeterminava i parametri sulla base dei quali l'algoritmo risolveva una specifica operazione. In altri termini, l'algoritmo si limitava a dedurre in maniera *deterministica* conseguenze da assiomi prefissati dal programmatore.

Con il potenziamento enorme della capacità dei computer di immagazzinare dati e della loro velocità di calcolo, si afferma un approccio "statistico-probabilistico". Gli algoritmi di nuova generazione, grazie all'istruzione a cui tutti noi contribuiamo inconsapevolmente (e gratuitamente), sono in grado di elaborare *autonomamente* criteri di inferenza e formulare proposizioni a partire da determinate premesse. Criteri che, in molti casi, risultano "incomprensibili" persino agli stessi programmatori. Nell'infosfera, i dati sono tutto e non v'è necessità di una teoria che spieghi la causalità della loro correlazione. É sufficiente la correlazione in sé tra grandi quantità di dati (i big data) a generare la conoscenza (Cantaro, 2024). L'idea che tali sistemi possano operare senza una supervisione o una "comprensione" umana del loro processo decisionale solleva domande di trasparenza e responsabilità.

La distinzione sempre più sfumata tra mezzo e soggetto

La libertà, cioè la capacità di determinarsi autonomamente rispetto alla propria corporeità ed alla propria storia è (almeno per ora) prerogativa esclusiva dell'uomo cui consegue l'assunzione della responsabilità per le proprie azioni. È ciò che è scritto con termini diversi nell'articolo 2 della Costituzione che riconosce i diritti inviolabili nel contesto della solidarietà.

Il rapporto tra libertà – responsabilità presuppone una *distinzione concettuale* che è posta alla base del diritto civile, penale, amministrativo: la distinzione tra "strumento" o mezzo e soggetto agente (Simoncini, 2019). Lo strumento è quell'elemento in un corso di azioni che contribuisce a causare un evento ma che non esercita alcuna autonomia. Il soggetto è invece colui che dà avvio a quel corso di azioni causando l'evento. In questo orizzonte concettuale, «il responsabile di un evento è chiunque abbia deciso il corso di azioni che ha causato l'evento stesso; ovviamente, la sua responsabilità (in senso morale e giuridico) si può graduare fino a scomparire, a seconda del grado di coscienza e volontà dell'evento stesso». Fino a qualche anno fa, nessuno avrebbe immaginato di ritenere responsabile della morte di un uomo un autoveicolo perché "tecnicamente" è stato il mezzo che, colpendo la persona, ne ha causato l'evento morte. Oggi, dinanzi ad auto senza guidatore questioni del genere stanno diventando molto più complesse da dipanare.

Un numero crescente di decisioni che incidono sulle libertà individuali e collettive sono "prese" o "supportate" da macchine ovvero da sistemi automatizzati governati da algoritmi: dai navigatori satellitari alla gestione del traffico aereo, dalla medicina alle decisioni giudiziarie. E il grado di supporto offerto da tali programmi non è sempre facile da distinguere dall'attività umana. Illustrerò tre casi paradigmatici concernenti rispettivamente il diritto penale, il diritto amministrativo e il diritto del lavoro.

La "giustizia predittiva" e la macchina di Kafka

Il primo caso attiene al campo controverso della c.d. "giustizia predittiva" (Zaccaria, 2021 e Garapon, Lassègue, 2021). Siamo negli Stati Uniti. COMPAS è un prodotto software che utilizza algoritmi di IA per stabilire l'entità della pena da parte di un giudice penale. Il software in questione non "produce" la decisione, ma indica sulla base di un algoritmo decisionale *proprietario*, il fattore di rischio ovvero la probabilità della ripetizione del reato (alta, media, bassa). Come, in seguito, gli stessi giudici della Corte Suprema del Wisconsin (*Loomis c. Wisconsin*, 2016) hanno chiarito, l'indicatore prodotto dal software è stato solo uno degli elementi presi in considerazione per la decisione, non l'unico.

Vari elementi, tuttavia, fanno ritenere che questa separazione tra decisione umana e apporto tecnologico sia più difficile da tracciare sul piano pratico. Da un lato, assume rilievo anche per la "macchina giudiziaria", secondo l'espressione

con cui il PNRR definisce la giustizia, l'imperativo dell'*efficienza* che prescrive lo smaltimento degli arretrati. Dall'altro lato, il giudice posto di fronte a cause che per alcuni aspetti sono *seriali*, cioè caratterizzate da elementi di fatto e diritto ricorrenti e sollecitato dagli "indicatori di performance" è indotto a servirsi dell'apporto della macchina. Viene alla mente il racconto di Kafka, "*Nella colonia penale*" in cui l'ufficiale che presiede al funzionamento di una macchina deputata ad eseguire condanne mortali a un certo punto dice: «fin qui si è reso ancora necessario del lavoro manuale, ma d'ora in poi l'apparecchio lavorerà da solo».

Conoscibilità e comprensibilità dell'algoritmo

Il secondo caso paradigmatico concerne la legge sulla c.d. "Buona scuola" (n. 107 del 2015). il Ministero dell'Istruzione si è trovato a dover gestire un numero rilevantissimo di richieste di assegnazione, ovvero di mobilità, da ordinare, selezionare e decidere sulla base della complessa regolamentazione normativa, delle condizioni dei richiedenti, nonché delle disponibilità delle sedi di servizio. Il Ministero, quindi, ha affidato la redazione delle graduatorie ad un software prodotto da una società privata per considerare tutte le complesse variabili di natura legale e fattuale. Un sindacato di insegnanti chiede di effettuare l'accesso all'algoritmo di assegnazione, sulla base dell'art. 22 della legge 241 del 1990. Il Ministero nega tale accesso e il sindacato impugna il diniego dinanzi al Tar Lazio

La sezione III-bis del Tar Lazio, pronunciandosi sulla questione con la sentenza n. 3768/2017 ha sancito che «non è conforme al vigente plesso normativo complessivo e ai dettami dell'art. 97 della Costituzione ai principi ad esso sottesi agli istituti di partecipazione procedimentale definiti [dalla] L. 7/8/1990, n. 241, all'obbligo di motivazione dei provvedimenti amministrativi sancito dall'art. 3, stessa legge, al principio ineludibile dell'interlocuzione personale intessuto nell'art. 6 della legge sul procedimento e a quello ad esso presupposto di istituzione della figura del responsabile del procedimento, affidare all'attivazione di meccanismi e sistemi informatici e al conseguente loro *impersonale* funzionamento, il dipanarsi di procedimenti amministrativi, sovente incidenti su interessi, se non diritti, di rilievo costituzionale».

Il Consiglio di Stato a sua volta, pur riconoscendo che «l'impiego di procedure informatiche nell'azione amministrativa è una scelta che deve essere "incoraggiata", in quanto rispondente ai princìpi generali di efficienza ed economicità (art. 1, l. 241/1990), ha ribadito categoricamente che si possono

impiegare tecnologie algoritmiche all'interno dei procedimenti amministrativi, ma imponendo per la loro utilizzabilità, una serie stringente di condizioni che garantiscano la piena *conoscibilità* e *comprensibilità* di tali algoritmi in tutte le loro componenti. Se, tuttavia, la conoscibilità è garantita dalla possibilità di accedere al codice sorgente del programma di IA, sia pure tramite l'aiuto di un professionista, molto più difficile da garantire è la comprensibilità dell'IA di ultima generazione persino per lo stesso programmatore.

L'algoritmo discriminatorio

Ammettiamo pure che l'algoritmo sia rispettoso del *principio di conoscibilità e comprensibilità* (principio tra l'altro desumibile dall'art. 42 della Carta dei diritti fondamentali dell'UE) e che il suo utilizzo non rappresenti la motivazione esclusiva della decisione giudiziale, l'algoritmo stesso potrebbe però essere di per sé *discriminatorio* e, dunque, incostituzionale. Questo accade, in primo luogo, quando l'algoritmo predittivo è costruito su un set di dati che è già in partenza discriminatorio. Ad esempio, alcuni ricercatori hanno mostrato che l'algoritmo COMPAS sovrastimava sistematicamente il rischio di recidiva per gli imputati neri ed altrettanto sistematicamente sottostima il rischio per i bianchi.

Il terzo caso paradigmatico concerne l'ipotesi di una discriminazione "indiretta" come tale più difficile da rivelare. Si tratta di una controversia nata tra alcune associazioni sindacali ed una azienda del settore del "delivery" per l'accertamento della natura discriminatoria di una "decisione" presa, appunto, dall'algoritmo impiegato. Siamo nel campo delle nuove forme di lavoro "autonomo" (nei fatti subordinato e iper-precario) tramite piattaforme digitali, meglio note come gig-economy". La distribuzione e la gestione dei flussi di lavoro avvengono attraverso una piattaforma digitale che individua le sessioni di lavoro. I rider devono scaricare l'applicazione sul cellulare e creare un proprio profilo per accedere alla piattaforma digitale. Il rider, una volta prenotata la sessione di lavoro, qualora non abbia provveduto a cancellarsi entro 24 ore antecedenti il turno, è tenuto obbligatoriamente a recarsi all'interno del perimetro della zona di lavoro prenotata al fine di loggarsi entro un massimo di 15 minuti dall'inizio del turno. Il mancato log-in entro 15 minuti o la cancellazione del turno a meno di 24 ore dall'inizio della sessione di lavoro, determinano una penalizzazione delle statistiche del rider.

Nel caso promosso dai Sindacati contro la piattaforma *Deliveroo* si denunciava

dinanzi al giudice del lavoro di Bologna che l'algoritmo produceva decisioni *discriminatorie* nei confronti di quei lavoratori che, dopo essersi prenotati, si astenevano dal lavoro, per partecipare ad uno sciopero, ovvero per malattia o per altra causa prevista dall'ordinamento.

Accoglieva il ricorso il Trib. Bologna, 31 dicembre 2020: «è proprio in questa incoscienza (…) e cecità (…) del programma di elaborazione delle statistiche di ciascun rider che alberga la potenzialità *discriminatoria* dello stesso»…«Il sistema di profilazione dei rider adottato dalla piattaforma (…), basato sui due parametri della affidabilità e della partecipazione, nel trattare nello stesso modo chi non partecipa alla sessione prenotata per futili motivi e chi non partecipa perché sta scioperando (o perché è malato, è portatore di un handicap, o assiste un soggetto portatore di handicap o un minore malato, ecc.) in concreto discrimina quest'ultimo, eventualmente emarginandolo dal gruppo prioritario e dunque riducendo significativamente le sue future occasioni di accesso al lavoro».

Secondo questa giurisprudenza, dunque, non basta che l'algoritmo utilizzato per decidere sia essere pienamente *conoscibile* e accessibile sia pure tramite l'aiuto di un professionista (fermo restando il problema della insondabilità dei programmi di IA basati sull'apprendimento automatico), ma la decisione prodotta dall'algoritmo dev'essere anche *ragionevole*, se parametrato ad una decisione umana.

Da questo punto di vista, risulta ingenua la raffigurazione dell'algoritmo e dell'intelligenza artificiale come garanzia di neutralità, imparzialità, correttezza nelle decisioni (Cantaro, 2024). Essi, infatti, sono il prodotto degli indirizzi forniti da parte di chi li detiene, sono finalizzati ai risultati per i quali vengono elaborati e cioè, nel caso delle piattaforme di lavoro, la *massimizzazione del profitto*.

La responsabilità civile dell'IA

L'ultimo profilo da prendere in considerazione è quello propriamente della *responsabilità civile per danni*, si pensi all'auto a guida autonoma, a un drone di consegna andato fuori controllo e così via (in generale Teubner, 2019).

Una parte della dottrina si è interrogata, sulla possibilità di considerare i robot, e comunque tutte le macchine guidate da IA come *(pseudo)-soggetti di diritto* e sull'eventuale necessità emergente di creare una nuova categoria, la cosiddetta *personalità elettronica* (Ruffolo, 2020). Altra dottrina ha correttamente rilevato

che la personalità elettronica è di difficile inquadramento sia come persona giuridica, poiché tale ipotesi implicherebbe l'esistenza di soggetti idonei a rappresentarla e dirigerla, sia come persona fisica, comportando tale ultima lettura il rischio etico di umanizzare eccessivamente i dispositivi guidati da IA. Allo stato attuale della tecnologia, le macchine non svolgono un'azione sufficientemente originale e indipendente tale da poterle considerare come centro autonomo di imputazione giuridica.

Maggiore consenso si riscontra sulla qualificazione della responsabilità da IA quale responsabilità *oggettiva*, intesa come responsabilità contrapposta a quella per colpa (Alpa 2020). In questo caso, sarebbe sufficiente dimostrare il danno e la connessione causale tra questo e il funzionamento che lo ha generato ed una complessa valutazione dei rischi e della capacità di controllo da parte di persona che, in determinate circostanze, è in grado di minimizzare i rischi e affrontare l'impatto negativo" dell'intelligenza artificiale.

Discussa è tuttavia la tipologia di responsabilità oggettiva. Vari argomenti, a seconda degli autori, militano a favore di una configurazione come ipotesi di responsabilità da attività pericolosa (art. 2049 c.c.) oppure come responsabilità da cosa in custodia (2051 c.c.) ed è stata proposta, persino, l'ipotesi della responsabilità da animale domestico (2052 c.c.). Tre fattispecie riconducibili al tema dell'IA solo al prezzo di forzature interpretative.

Dal punto di vista del diritto dell'Unione europea, è stata proposta l'applicazione della responsabilità da prodotto difettoso – Direttiva 85/374/CEE, confluita nel Codice del Consumo (D.lgs. 206/2005) – sebbene l'imprevedibilità e l'eterogeneità dei modelli di AI sembri sfuggire alle definizioni di prodotto e difetto presupposti della disciplina. Abbastanza forzata pare anche l'interpretazione estensiva di tale ultima ipotesi di responsabilità al programmatore dell'algoritmo che "governava" il prodotto. Altre difficoltà nell'attribuzione della responsabilità possono derivare dal fatto che alcuni sistemi di IA sono in grado di modificarsi in autonomia in base all'elaborazione di nuovi dati (*self-adapting*), oppure dagli aggiornamenti cui sono soggetti, o dalla loro interazione continua con l'ambiente circostante, altri sistemi e fonti di dati.

Da ultimo, in tempi recenti la Commissione ha avanzato una proposta di direttiva sulla responsabilità per l'intelligenza artificiale. La proposta normativa, per un verso, alleggerisce l'onere della prova a carico dei danneggiati introducendo la

“presunzione di causalità” – appunto un caso di responsabilità oggettiva. Per l’altro la direttiva mira ad aiutare i danneggiati ad accedere a elementi di prova pertinenti. I danneggiati potranno chiedere al giudice di ordinare la divulgazione di informazioni sui sistemi di IA ad alto rischio.

Gli irrisolti interrogativi sulla giustizia predittiva, sulla conoscibilità (e comprensibilità) degli algoritmi, sulla loro origine discriminatoria e ancora sulla responsabilità civile del danno causato dall’IA confermano che «anche in presenza dei più perfezionati algoritmi si è obbligati a rimandare a qualcosa di *esterno* al loro meccanismo, a una responsabilità e a una libertà radicale che forse non esiste neppure, e che coincide infine con quella essenziale incompletezza che la tradizione filosofica e sapienziale, come pure le ricerche sulla natura della coscienza, hanno ontologicamente identificato come *l’essenza stessa dell’uomo*» (Zellini, 2018).

Testi di riferimento

Alpa G., *Diritto e intelligenza artificiale. Profili generali, soggetti, contratti, responsabilità civile, diritto bancario e finanziario, processo civile*, Pacini Editore, 2020.

Cantaro A., *Vita e lavoro nel tempo dell’IA*, in https://fuoricollana.it/vita-e-lavoro-nel-tempo-dellia/.

Garapon, A., Lassègue, J., *La giustizia digitale. Determinismo tecnologico e libertà*, trad. it., Il Mulino, Bologna, 2021.

Harari Y.N., *Nexus. Breve storia delle reti di informazione dall’età della pietra all’IA*, Bompiani, 2024.

Ruffolo, *La “personalità elettronica*”, in *Intelligenza artificiale: il diritto, i diritti, l’etica*, Giuffrè, 2020, 214 ss.

Simoncini, A. 2019. *L’algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, *BioLaw Journal – Rivista di BioDiritto*, 1, pp. 63-89.

Teubner G., Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi, Edizioni Scientifiche Italiane, Napoli 2019.

Zaccaria, G. 2021. *Mutazioni del diritto: innovazione tecnologica e applicazioni*

predittive, in *Ars Interpretandi*, X/2021, n. 1, pp. 29-52.

Zellini, P. 2018. *La dittatura del calcolo*, Adelphi, Milano.