Antonio Pescapè

UN "NUOVO ORDINE" PER LA GIUSTIZIA?
Una riflessione tra algoritmi e diritto*

Quando si discute *lato sensu* di Intelligenza Artificiale oggi, lo si fa analizzando soprattutto specifiche aree tematiche quali *Computer Vision, Pattern Recognition, Automated Reasoning, Game theory, Logics, Multi-Agents, Fuzzy systems, Knowledge representation, Speech Recognition, Natural Language Processing, Machine Learning, Deep Learning, Cognitive Robotics*. Tutti questi ambiti sono stati oggetto di specifici lavori su riviste internazionali e sono stati alla base di una grande quantità di conferenze internazionali di settore, ne consegue che si può facilmente comprendere come il termine Intelligenza Artificiale sia un termine "*umbrella*" sotto il quale vanno a collocarsi oggi ambiti molto diversi tra loro.

Bisogna anche aggiungere un'altra importante distinzione, quella tra *Strong AI* e *Weak AI*; con *Strong AI* ci si riferisce a scenari nei quali le macchine sono effettivamente in grado di pensare ed eseguire compiti da sole, essere veri e propri *alias* di un essere umano. In questa categoria sono ricompresi il *Turing Test* (Turing), il *Coffee Test* (Wozniak), il *Robot College Student Test* (Goertzel), l'*Employment Test* (Nilsson), il *Flat Pack Furniture test* (Tony Severyns), il *Mirror Test* (Tanvir Zawad). Con *Weak AI* ci si riferisce invece a scenari in cui ci si concentra su un compito ristretto, riferendosi al fenomeno per il quale le macchine che non sono "troppo intelligenti" per svolgere il proprio lavoro possono essere costruite in modo tale da "*sembrare* intelligenti", simulando esclusivamente la funzione cognitiva umana. Questo tipo di AI, definita "debole" agisce in

---

* Questo lavoro nasce dall'intervento tenuto dall'autore nell'ambito della Tavola Rotonda su "Intelligenza artificiale nella decisione politica" ospitata all'interno del Convegno annuale del CIRB "Il tempo dell'umano e il tempo delle macchine" tenutosi a Villa Doria D'Angri, Università degli Studi Parthenope, via Petrarca 80, Napoli presso il 26 e 27 novembre 2021. L'intervento è stato leggermente ampliato in fase di riscrittura al fine di rendere più chiare alcune definizioni. Non ha certamente la pretesa di esaustività, ma si pone come spunto di riflessione su aspetti e problemi legati alle nuove prospettive di una giustizia sempre più "algoritmica".

maniera "semplice" ed è vincolata dalle regole che le vengono imposte senza mai poter andare oltre queste regole.

Oltre queste due definizioni di *Strong AI* e *Weak AI*, la tassonomia dell'Intelligenza Artificiale si specializza ulteriormente in altri quattro tipi, quella di tipo 1: si riferisce a una AI reattiva e specializzata in una singola area. Ad esempio, la redazione e la revisione di contratti di finanziamento di natura commerciale. Quella di Tipo 2 riguarda una AI che ha memoria o "esperienza" appena sufficiente per prendere decisioni adeguate ed eseguire azioni appropriate in situazioni o contesti specifici; a quella di Tipo 3 corrisponde una AI che ha la capacità di comprendere pensieri ed emozioni che influenzano il comportamento umano. Infine (per ora) l'AI di Tipo 4 che si configura come quella AI più vicina a quelle tipicamente rappresentata nelle rappresentazioni cinematografiche/nella fantascienza: macchine autocoscienti, super-intelligenti e senzienti.

Questa dicotomia tra AI cognitivo-produttiva e ingegneristico-riproduttiva è superata dalle linee di ricerca più recenti che si approcciano ad una AI "indipendente" con capacità di agire autonoma e non una mera riproduzione della intelligenza umana. L'intuizione di fondo è quella che l'AI vada alimentata con nozioni elementari, come nel caso di un bambino, che agisce, o prova ad agire, non perché sappia fare bensì perché osserva qualcuno (un adulto, ad esempio) e prova, tenta, a ripetere (riprodurre) quella data cosa o una simile. Questa intuizione negli approcci di *Weak AI* si sostanzia nella fase di apprendimento (*training*) – durante la quale si costruisce il modello addestrando una rete – e in quella di uso (*testing*) – durante la quale si valuta la capacità del sistema addestrato di operare su dati diversi da quelli utilizzati nella fase di addestramento. Il *Machine Learning* (ML) – e il più recente *Deep Learning* (DL) – fa parte di questa categoria. Sebbene approcci di *machine learning* fossero stati già proposti alla fine degli anni '50, l'*hype* a cui si sta assistendo è legato alla combinazione di tre importanti condizioni: disponibilità di considerevole potenza di calcolo a costi contenuti, disponibilità di considerevole spazio di memorizzazione a costi contenuti (entrambi abilitati anche dal paradigma cloud), infine disponibilità di quantità di dati mai viste prime. Grazie a queste tre condizioni, gli algoritmi di ML e di DL oggi cominciano ad avere performance tali da poter prevedere un loro utilizzo massivo in ogni attività quotidianamente.

Proprio questo utilizzo massivo di algoritmi di *Weak AI* pone una serie di questioni complesse che oggi rappresentano la frontiera della ricerca in questo settore. Innanzitutto, l'AI deve avere una serie di proprietà: deve essere etica, trasparente, affidabile, antropocentrica, inclusiva, responsabile e neutrale.

Ma per quanto detto in apertura è chiaro che i *bias* introdotti dai dati di addestramento rischiano, anzi certamente condizionano l'AI, "orientandone" il comportamento. I sistemi di apprendimento dell'AI sono alimentati, addestrati e interpretati da esseri umani e quindi potenzialmente pieni di pregiudizi, sia consci che inconsci: diversi sono i casi che dimostrano questa complessità. È quello che è accaduto a Twitter accusato di razzismo dopo che i suoi algoritmi (di ritaglio delle foto) "preferivano" i volti più chiari a quelli più scuri, anche Amazon è stato accusato di sessismo dopo aver utilizzato uno strumento di intelligenza artificiale per lo smistamento dei CV, che aveva imparato a favorire i candidati uomini, c'è stato poi il caso di una Corte americana che ha utilizzato un algoritmo di AI addestrato su dati relativi a reati di una particolare area di Los Angeles e nel fare *screening* è stato scoperto che l'algoritmo "soffriva" di un "*bias* di razza", e che i reati erano commessi esclusivamente da afroamericani.

Bisogna aggiungere la questione della interpretabilità e fino a quando non si sarà in grado di comprendere perfettamente i meccanismi e le motivazioni per le quali un algoritmo di AI prende una decisione, non potrà mai essere utilizzato in contesti dove il risultato (responso, verdetto) richiede una "certificazione" frutto di una interpretabilità del processo di decisione.

Molte sono le preoccupazioni nei confronti dell'AI, ma sono molti anche i tecnoentusiasti. Già 2019, la Commissione europea ha pubblicato le "Regole etiche per un'AI affidabile", sancendo la necessità di sostenere lo sviluppo e l'adozione di un'AI etica e affidabile in tutti i settori, a condizione che sia "etica, sostenibile, incentrata sull'uomo e rispettosa dei diritti e dei valori fondamentali". È sempre più essenziale comprendere e misurare la correttezza dell'AI, e ciò può essere fatto in diversi modi, come richiedere che i modelli abbiano un valore predittivo medio uguale per tutti i gruppi, o che i modelli abbiano tassi di falsi positivi e falsi negativi uguali per tutti i gruppi. In particolare, la nozione di "equità controfattuale" considera una decisione equa per un individuo se è la stessa nel mondo reale rispetto a un mondo alternativo in cui l'individuo apparterrebbe a un gruppo demografico diverso.

La soluzione è chiaramente nella combinazione di AI e umano ma, nonostante le non poche perplessità, se l'AI viene implementata correttamente può svolgere un ruolo decisivo per lo sviluppo e può prendere decisioni più eque e obiettive rispetto a quelle prese da un (solo) essere umano.

L'intelligenza artificiale è essenziale per ottenere informazioni preziose da dati su larga scala, tuttavia, è necessario prestare attenzione all'implementazione e all'addestramento corretto, ma anche considerare, come si

è già visto, i *bias* che si nascondono nei dati e piuttosto che accusare l'AI per i pregiudizi, dovremmo considerare più da vicino il fattore umano e imparare a gestire l'Intelligenza Artificiale Generale. A tal proposito, il Regolamento Europeo sull'AI dell'aprile 2021, giunto piuttosto in ritardo, propone una classificazione delle applicazioni di AI: Vietate, Alto Rischio, Rischio Limitato, Rischio Minimo. Nella classe delle applicazioni Vietate troviamo: l'uso di sistemi di AI che distorcono il comportamento di una persona attraverso tecniche subliminali, l'uso di sistemi di AI che sfruttano qualsiasi vulnerabilità in modo da causare o essere suscettibili di causare danni fisici o psicologici, l'uso di sistemi di IA che consentono la valutazione/classificazione dell'affidabilità di persone fisiche mediante l'attribuzione di un punteggio sociale (*social score*).

A questo punto è necessario porsi un quesito: il sistema predittivo abilitato dall'AI che influenza gli esseri umani cambiando il loro processo decisionale e, di conseguenza, il loro comportamento è accettabile? È necessario tener conto che il sistema predittivo diviene poi informazione predittiva consentendo così di moltiplicare la capacità di fare propaganda (ne abbiamo avuto contezza con il COVID-19, e adesso sta accadendo nella guerra Russa/Ucraina). La novità è che oggi si può mirare in maniera sempre più raffinata agli individui, generando una informazione personalizzata che è, probabilmente, la più grande minaccia per la stabilità della società e della democrazia, per come oggi la si conosce.

Gli algoritmi di AI sono ufficialmente in uso nei tribunali francesi dal 2020, in piena emergenza sanitaria con decreto n. 2020/356 del 27 marzo 2020, è stato autorizzato «DataJust», un trattamento automatizzato dei dati personali con cui si mira a sviluppare, per un periodo di due anni, un dispositivo algoritmico che permetta di identificare gli importi richiesti e offerti dalle parti di una controversia e gli importi assegnati alle vittime a titolo di risarcimento per i danni alla persona nelle sentenze emesse in appello dai tribunali amministrativi e dai tribunali civili. Il sistema si basa sull›estrazione e l›elaborazione automatica dei dati contenuti nelle decisioni giudiziarie. L'applicazione di questo strumento è stata possibile grazie alla *Loi pour une République numérique* del 7 ottobre 2016, che ha autorizzato la pubblicazione di dati aperti di decisioni giudiziarie anonimizzate, a cui si aggiunta la *loi de programmation et de réforme pour la justice* del 23 marzo 2019 e il decreto del giugno 2020 che ha apportato ulteriori precisazioni come le condizioni per la messa a disposizione del pubblico delle decisioni giudiziarie (in particolare, i termini per la messa online, il diritto di accesso e di rettifica), il rafforzamento dell'anonimato che deve comprendere anche elementi che

consentano l'identificazione delle parti, in caso di rischio di violazione "della sicurezza o della *privacy* di queste persone o del loro *entourage*" (e non solo il nome e il cognome delle persone), il rilascio di copie a terzi, il calendario di diffusione. La diffusione è scaglionata fino al 2025: le sentenze della Corte di Cassazione già a settembre 2021, poi le decisioni civili, sociali e commerciali delle corti d'appello nella prima metà del 2022 (un flusso di 230.000 decisioni ad aprile 2022), seguite dalle sentenze dei tribunali amministrativi (giugno 2022) e degli altri tribunali, in particolare in materia penale. Tutto questo per consentire una migliore amministrazione della giustizia e la messa a disposizione dei singoli di uno strumento che consenta loro di effettuare scelte più informate circa l'opportunità o meno di avviare un contenzioso o di accettare o meno le offerte di risarcimento.

Si tratta di quasi 3,9 milioni di decisioni giudiziarie; per il momento è stata data priorità alle sentenze delle corti d'appello, in attesa della creazione di un portale (*Portalis,* che porta il nome del celebre giurista incaricato di redigere il Codice napoleonico) che diffonderà le decisioni dei tribunali di primo grado.

L'uso di questi strumenti sta cambiando profondamente il lavoro dei giudici, dei funzionari di giustizia e il mondo dell'avvocatura. Il loro sviluppo futuro solleva, come già si è potuto osservare in altri ambiti, non poche questioni etiche; non poche sono le perplessità e le preoccupazioni per possibili abusi che possono nascere in particolare nell'ambito del predittivo. Non sono pochi i timori circa l'applicazione di un sistema giudiziario automatico e "disumanizzato" e sempre più sentite sono le critiche nei confronti dell'applicazione dell'AI al verdetto, in diversi Stati sono già in corso esperimenti che prevedono l'utilizzo di software per la gestione della giustizia, alleggerendo l'attività dei tribunali e riducendo i costi. In Ontario (Canada), un "tribunale virtuale" è responsabile della risoluzione delle controversie tra vicini o tra dipendenti e datori di lavoro. In Quebec, il *software* viene utilizzato anche per risolvere piccole controversie commerciali. In Estonia, un robot dovrebbe presto stabilire la colpevolezza di una persona nelle controversie "minori" (meno di 7.000 euro).

È possibile che si sia davanti alla creazione di piattaforme per il diritto, come quelle del *marketplace*, d'altronde alcuni sistemi giudiziari (guardiamo il caso dei paesi di *common law)* dove la decisione è frutto soprattutto di un "precedente", ben si prestano alla giustizia algoritmica" ma, forse in Paesi dove la civiltà giuridica si fonda su una diversa e secolare tradizione della norma e della legge, potrebbe provocare un vero e proprio isterili-

mento culturale e un minor margine di manovra da parte degli attori del diritto. Ma siamo dinanzi ad un cambiamento epocale, non diverso da quello che si trovarono ad affrontare i giuristi all'indomani dell'applicazione delle norme codificate, sarà necessario, così, un lungo periodo di adattamento alla trasformazione, che porterà speriamo ad un uso più consapevole dei sistemi di AI.