

[as] intersezioni

Bella senz'anima

di Francesco Bianchini

filosofo della scienza



Non è facile oggi parlare di etica dell'intelligenza artificiale (IA). Non è facile come lo sarebbe stato qualche anno fa o, ancora di più, qualche decennio fa. Le ragioni di questa difficoltà sono varie. Innanzitutto, l'IA è una disciplina scientifico-tecnologica e l'etica è una materia filosofica. E fin qui le cose sembrano essere piuttosto chiare. Dopodiché, la congiunzione fra queste due discipline può avvenire, anche qui senza troppe complicazioni, come una riflessione filosofica sulla scienza che chiama in causa aspetti morali e valoriali. Come per molte altre discipline scientifiche, è possibile analizzare gli aspetti etici riguardanti i principi di una disciplina scientifico-tecnologica, le implicazioni morali delle sue metodologie e dei suoi risultati, e le conseguenze che il suo impatto può avere sulla società in cui essa viene sviluppata e diffusa. Si pensi alla fisica e alle

riflessioni sulla visione dello spaziotempo portate dalla relatività; alla genetica e agli interrogativi che suscita in quanto disciplina in grado di influenzare e modificare la costituzione del mondo vivente, compreso quello umano; alla fisica delle particelle e alle considerazioni etiche sull'uso, bellico o meno, dell'energia nucleare; alla medicina e ai tanti risvolti etici che le sue pratiche implicano. Fin qui il discorso non presenta difficoltà: una disciplina scientifica, in genere nei suoi aspetti applicativi, presenta quesiti morali che la filosofia della scienza e la filosofia morale sono abituate ad affrontare. Il discorso per l'IA si complica se la consideriamo nelle sue varie sfaccettature. Si possono individuare due macroaree entro cui si sviluppa la riflessione etica relativa all'IA.

La prima riguarda, nel senso più classico relativo all'etica di una disciplina scientifica, questioni legate

a.
L'intelligenza artificiale alle prese con le scelte etiche.



b.
L'IA consente nuove opportunità nel mondo del lavoro, semplificando i processi e liberando tempo per attività a maggiore valore umano.

al suo sviluppo, alla sua applicazione e ai suoi risultati. Questa area è molto vasta e va da aspetti sociali e politici dell'impiego dell'IA a questioni più strettamente relative all'uso, diretto o indiretto, dell'IA da parte del singolo o in relazione all'individuo e alle sue attività. Questioni rilevanti da tale punto di vista riguardano l'impatto che l'IA ha sulla nostra società, sul mondo del lavoro, sulle *policy* da prevedere per un corretto uso degli strumenti IA disponibili.

L'IA consente nuove opportunità nel mondo del lavoro, semplificando i processi e liberando tempo per attività a maggiore valore umano.

Alcune delle domande cui si vuole rispondere sono: quale peso può avere l'IA nel determinare le scelte politiche e sociali? Come può influenzare i sistemi politici democratici, ad esempio per quanto riguarda i loro meccanismi elettivi e di scelta? Quale impatto ha l'IA sulle dinamiche occupazionali? Porterà a cambiamenti che cancelleranno posti di lavoro con terribili conseguenze sociali? O ne creerà di nuovi, dimostrandosi un'opportunità per le evoluzioni future delle società in cui si diffonderà capillarmente? O entrambe le cose? E in tal caso, con che tempistiche? Quali linee guida e sistemi di leggi devono essere previsti e strutturati per prevenire usi pericolosi dell'IA, sia per le società che per gli individui, e quali agenzie nazionali e transnazionali sono titolate e proporli? A tale gamma di preoccupazioni si sommano, infine, le questioni legate alla sostenibilità ambientale ed energetica dell'IA con il loro conseguente impatto sociale.

Inoltre, come si diceva, i problemi etici sociali si intrecciano con quelli relativi agli individui. Due fattori hanno reso pressante il problema etico dell'IA nei confronti delle singole persone: la tipologia algoritmica largamente prevalente dell'IA contemporanea e la sua recente diffusione su ampia scala. È vero infatti che oggi, anche più che soltanto dieci anni fa, l'IA è accessibile a un largo pubblico e pertanto è utilizzata, in senso diretto o indiretto, da utenti non specializzati. Ciò può generare usi incontrollati o poco consapevoli, come ad esempio la produzione di testi o immagini attraverso IA generative senza le opportune cautele e la necessaria competenza; o anche l'utilizzo dell'IA per manipolare/controllare gli individui, ad esempio attraverso algoritmi di profilazione in campo commerciale, finanziario, sociale, ecc. Inoltre, tali algoritmi sono in genere basati su reti neurali, in particolare sulle loro evoluzioni degli ultimi quindici anni, cioè vale a dire su metodi statistico-predittivi in grado di produrre risultati con un elevato grado di accuratezza, ma processualmente opachi. La loro elaborazione consiste in complesse funzioni matematiche che associano un

output a un input senza che i processi interni dell'esecuzione dell'algoritmo siano spiegabili in termini comprensibili, senza che sia possibile, cioè, assegnare loro un significato. La riflessione etica in questo caso si concentra, a partire da vari tentativi per raggiungerne la spiegabilità, su come proteggere la *privacy* degli utenti, sulla trasparenza dell'operato di tali sistemi, sul corretto modo di addestrarli partendo da dati che non siano condizionati da assunzioni implicite ingiuste o scorrette da un punto di vista etico e sociale (i cosiddetti *bias*). Negli ultimi anni sono state approntate numerose linee guida etiche, che condividono in genere principi quali trasparenza, equità, giustizia, beneficenza, non maleficenza. La seconda macroarea etica relativa all'IA, che la rende più originale rispetto ad altri campi scientifico-tecnologici, riguarda l'etica che i sistemi stessi di IA devono possedere. Si tratta della questione delle "macchine morali" o dell'"etica delle macchine". Poiché una delle caratteristiche che accomuna i sistemi dell'IA è la loro autonomia, essa deve essere in qualche modo normata anche dall'interno. Quali regole devono

guidare il comportamento delle IA? Quali comportamenti sono permessi e quali no? Ciò vale per i sistemi software, ma è ancora più evidente per i sistemi robotici; e vale sia per gli ambienti strutturati (si pensi ad esempio alla robotica industriale o a quella chirurgica) sia per quelli più indeterminati, dove è maggiore il grado di autonomia e di azione "libera" che i sistemi possono mettere in atto, specialmente quando sono presenti esseri umani (si pensi ad esempio al caso dei veicoli autonomi o a quello dei robot assistivi). È proprio l'interazione con gli esseri umani che fa nascere la richiesta di regole e norme per un comportamento adeguato e non dannoso, di una sorta di morale del sistema IA, anche nei casi estremi come quelli delle armi autonome. Seppure non sempre gli aspetti etici dell'IA vengano perseguiti o tenuti in debita considerazione, essi sono pervasivi e richiedono, oggi più che mai in misura crescente, competenze giuridiche, etiche ed epistemologiche con l'obiettivo morale effettivo di una IA non tanto migliore quanto più sostenibile per l'essere umano.

C.

Un "medico robot" umanoide che prepara la cartella medica in ospedale.

