DOPPIOZERO

Può l’intelligenza artificiale essere etica?

Tiziano Bonini

16 Maggio 2022

Recentemente sono usciti alcuni libri molto importanti per la comprensione dell’impatto dell’intelligenza artificiale sulla società. Il primo, di cui avevo scritto qui, è quello di Kate Crawford, *Né artificiale né intelligente. Il lato oscuro dell’IA* (Il Mulino, 2021, 312 pp., versione italiana dell’originale inglese *Atlas of AI. Power, Politics and the Planetary Costs of Artificial Intelligence,* Yale University Press, 2021). Il secondo è quello di Luciano Floridi, *Etica dell’intelligenza artificiale* (Raffaello Cortina Editore, 2022, 384 pp.). Il primo rappresenta la *pars destruens* del discorso pubblico sull’IA, il secondo, invece, la *pars costruens*.

Il libro di Floridi è altrettanto ambizioso, e altrettanto informato da anni di ricerca del suo gruppo di lavoro e dall’analisi di una mole notevole di letteratura sull’IA.

Crawford considera l’IA come un “apparato socio-tecnico”, cioè un insieme di attori, sia umani che non umani, istituzioni, governi, tecnologie in relazione tra loro. Questo tipo di approccio, molto noto tra gli studiosi di STS (*Science and Technology Studies*), è anche molto vicino a quello di Bruno Latour, sociologo francese tra i fondatori dell’*Actor Network Theory*. In estrema sintesi, per descrivere il pensiero di Latour sulla tecnologia potremmo citare il titolo di un suo articolo molto famoso, “technology is society made durable” (1990): la tecnologia è una specie di “calco” durevole della società, cioè, porta inscritte su di sé, nei suoi aspetti tecnici, le norme e i valori della società da cui è emersa. Secondo questo approccio, la tecnologia non determina la società, non ha qualità intrinseche che la rendono buona o cattiva, ma non è neanche neutrale, né può essere resa “buona” o “cattiva” a piacimento. La tecnologia è sempre incastrata in una rete di relazioni tra attori di diverso genere, che ne modellano e ne stabiliscono l’evoluzione. In questa visione, tecnologia e società si modellano e si costituiscono reciprocamente. La Crawford, influenzata da questo approccio, riesce a dare conto degli attori (istituzioni, governi, scienziati, aziende, innovazioni tecnologiche) che hanno fin qui modellato il nostro uso dell’IA.

Se Crawford è una sociologa della scienza e della tecnologia, Floridi è un filosofo, poco interessato alle dinamiche sociali.

Al contrario di Crawford, egli non descrive mai l'IA come un apparato socio-tecnico, cioè non prende in considerazione ciò che avviene "intorno" all'intelligenza artificiale, qual è la nuvola di attori dai quali sono emersi gli usi attuali dell'IA. Per l'autore italiano, l'IA è "solo" una serie di innovazioni tecnologiche.

Il libro è molto ben orchestrato e ambisce ad essere divulgativo, nonostante alcuni capitoli molto tecnici. Si apre con una breve ricostruzione storica dei processi di digitalizzazione della società, che però di storico ha molto poco e riproduce una visione molto lineare di questo processo: al progressivo aumento della potenza di calcolo, fa corrispondere un progressivo stato di digitalizzazione della società, fino ad arrivare alla commistione tra mondo offline e online, che lui notoriamente chiama *onlife*. Poi passa a descrivere le diverse ondate di dibattito pubblico che hanno caratterizzato l'IA, fino a quella presente, e mette in evidenza le alterne fortune che la ricerca sull'IA ha avuto dalle sue origini fino ad oggi. Infine, si lancia sulle previsioni sul futuro dell'IA, legando questo futuro a questioni etiche, che sono il fulcro della seconda parte del libro e anche la disciplina in cui la ricerca di Floridi è più solida.

Di fatto, il libro si pone delle domande importantissime relative alla possibilità di far funzionare l'IA in maniera etica. Il valore del libro non sta nelle domande che pone, ma nelle risposte che prova a dare.

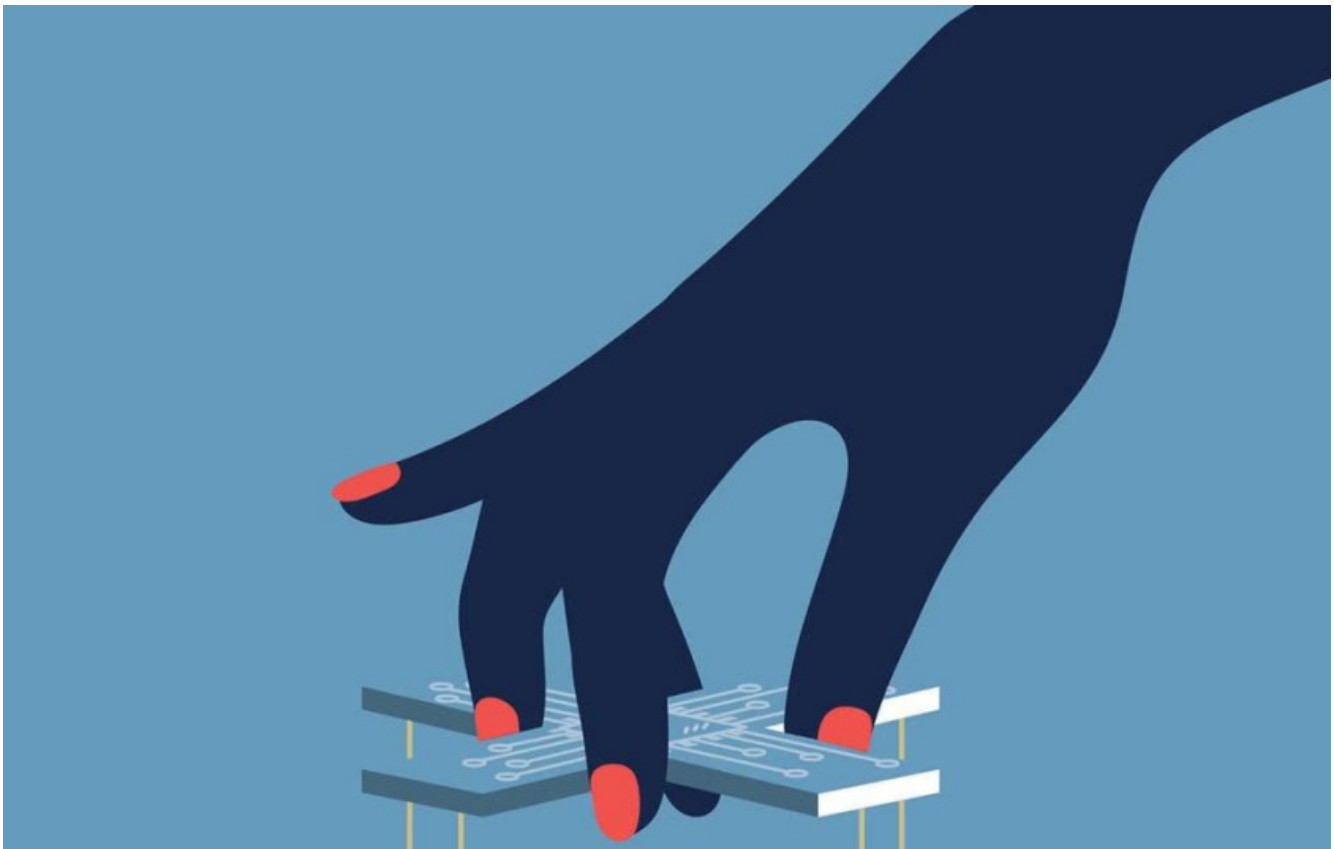
All'inizio della seconda parte del libro (capitolo 4) Floridi analizza in maniera comparativa tutti i principi etici sviluppati finora, da diversi enti, istituzioni e governi, per l'adozione di una IA socialmente vantaggiosa e propone una sintesi di questi principi, riducendoli a cinque macro-principi fondamentali, i primi quattro dei quali sono comuni a quelli della bioetica: *beneficenza, non maleficenza, autonomia* e *giustizia*. A questi quattro principi della bio-etica, Floridi aggiunge il principio dell'*esplicabilità*: cioè, l'IA, per essere definita "etica", deve essere intelligibile e responsabile (*accountability*), il suo funzionamento deve cioè essere "spiegabile".

Questa prima mappatura dei principi di base ai quali si dovrebbe ispirare qualsiasi implementazione di IA pone le fondamenta del discorso successivo. Floridi è molto consapevole delle conseguenze non etiche (discriminatorie) dell'impiego di algoritmi di IA e ne fornisce un'ampia descrizione nel capitolo 7. Qui è molto chiaro nel sostenere che l'IA, anche quando viene impiegata con buone intenzioni, può finire per riprodurre o amplificare discriminazioni già presenti nella società, come ormai molti studi e libri hanno dimostrato (Safiya Noble, *Algorithms of Oppression*, 2018; Virginia Eubanks, *Automating Inequality*, 2018; Kathy O'Neil, *Armi di distruzione matematica*, 2016).

Nel capitolo 8 invece analizza i possibili usi "impropri" e "illeciti" dell'IA. Tra le cattive pratiche inserisce i modi in cui l'IA può essere impiegata per generare profitti illeciti tramite azioni criminali come furti, frodi informatiche, reati contro la persona (comprese le molestie e il razzismo perpetrati da profili di bot sui social media), reati finanziari.

Dopo aver passato in rassegna i modi in cui l'IA può essere messa al servizio di atti criminali, Floridi passa all'analisi delle buone pratiche, che chiama AI4SG (*AI for Social Good*, "AI per il bene sociale") e identifica sette fattori etici essenziali per le future iniziative di AI4SG.

Per AI4SG Floridi intende "il design, lo sviluppo, l'implementazione di sistemi di IA in modo da (i) prevenire, mitigare, risolvere i problemi che incidono negativamente sulla vita umana e/o sul benessere del mondo naturale e/o (ii) consentire sviluppi preferibili dal punto di vista sociale e/o sostenibili dal punto di vista ambientale" (p. 227).



KATE CRAWFORD

NÉ INTELLIGENTE NÉ ARTIFICIALE

IL LATO OSCURO DELL'IA

il Mulino

A questo proposito, Floridi fornisce sette iniziative di successo in cui l'IA è stata già implementata per il "bene sociale". Tutte queste iniziative rispondono ai cinque principi etici che dovrebbero guidare l'IA, individuati nel capitolo 4.

Qui emerge finalmente uno dei tratti distintivi di questo libro, cioè la volontà di Floridi di non fermarsi solo alla critica degli aspetti problematici dell'IA (capitoli 7 e 8), ma arrivare a definire dei principi etici per il suo sfruttamento positivo.

A partire dal capitolo 9 fino alla fine, Floridi cerca razionalmente di individuare principi e criteri che dovrebbero informare il design e la governance degli algoritmi per perseguire apertamente "il bene sociale". Il fine ultimo di questo libro è quello di giungere a una definizione di un'etica dell'intelligenza artificiale (che Floridi chiama "etica del digitale") capace di individuare la maniera "migliore" e le azioni più "buone" per ottenere il massimo dei benefici dall'intelligenza artificiale e, in ultima istanza, "avere una società migliore".

Questa etica permetterebbe, secondo Floridi, di assimilare i potenziali benefici dell'IA, mitigandone i danni e i rischi.

Floridi, in sostanza, crede che sia possibile mettere l'AI al servizio della società, crede nell'esistenza di una IA "buona" e la contrappone ad una IA "cattiva", cosa alla quale la Crawford non crede affatto.

Come valutare, in conclusione, il libro di Floridi? Sarebbe scorretto situare Floridi tra i tecno-ottimisti. Egli non è un ingenuo "soluzionista tecnologico" che crede nella capacità della tecnologia di risolvere qualsiasi problema sociale. Infatti, rileva che "ci sono molte circostanze in cui l'IA non è il modo più efficace per affrontare un determinato problema sociale" (p. 225). In questo modo prende le distanze da facili accuse di "determinismo tecnologico". Nel cap. 10 infatti riconosce l'inutilità della classica diatriba tra apocalittici e integrati (che lui chiama, rispettivamente, "atei dell'IA" e "adepti della singolarità") e riconosce anche, come aveva fatto Crawford, che l'IA non è "intelligente", e difficilmente riuscirà a superare l'umanità nella capacità di "pensare", almeno nel breve-medio termine. È più corretto, infatti, comprendere l'IA come una tecnologia capace di risolvere problemi, portare a termine dei compiti, fare alcuni compiti meglio di come le facciamo noi, ma senza esserne "consapevole", cioè "senza pensare, limitandosi a elaborare una grande quantità di dati in modo sempre più efficace" (p. 275).

L’IA è come una lavastoviglie, ci dice Floridi, lava i piatti meglio di noi, ma non è più intelligente di noi. Da questo esempio, Floridi, un po’ ingenuamente, arriva a sostenere che “qualsiasi visione apocalittica dell’IA può essere ignorata” e che il rischio reale dell’IA sta nel fatto che “possiamo utilizzare male le nostre tecnologie digitali, a danno di una grande percentuale dell’umanità e dell’intero pianeta” (p. 276). In questo passaggio Floridi cade nella trappola della nota posizione retorica per cui la tecnologia non sarebbe “né cattiva né buona, dipende da come la usiamo”.

È su questa linea di faglia che si separano i libri di Crawford e Floridi. Quest’ultimo sembra valutare l’IA da un punto di vista puramente razionale ed etico, inserendolo in uno spazio vuoto, come se non esistessero la storia, l’economia, la politica e la cultura: come se, in sostanza, non esistesse una società che ha dato vita a questa forma attuale di IA. Possiamo concordare con Floridi sulla sua lista di norme etiche e attente all’impatto ambientale che dovrebbero guidare la progettazione di IA, ma queste norme etiche non si auto-impongono da sole, sono inserite in un contesto sociale, economico e politico che finora le ha rifiutate, o le ha sotto-stimate, per continuare ad estrarre profitto dai dati o per sorvegliare meglio milioni di persone.

Nelle 384 pagine del libro di Floridi non compare mai la parola “capitalismo” e solo 3 volte compare la parola “potere” intesa come “potere economico o politico”.

L’analisi di Floridi non mette in discussione il sistema politico ed economico che ha permesso, finora, di applicare l’IA in maniera non etica, non si interroga sulle cause sociali, politiche ed economiche di un uso non etico dell’IA. Nel suo resoconto, gli usi non etici dell’IA sono ascrivibili solo ai criminali, eppure i software di riconoscimento facciale basati su IA che discriminano i volti degli afro-americani e si comportano quindi in maniera non etica, non hanno finalità criminali. Eppure, nonostante questo, producono discriminazioni.

Alla fine del libro Floridi sostiene che il problema siamo noi, non la tecnologia, cioè dovremmo essere noi a rendere l’IA adatta al nostro ambiente, e non viceversa. L’IA dovrebbe essere usata per trattare le persone come fini, e non come semplici mezzi, ci dice ancora Floridi. Tutto molto condivisibile, certo, ma

dimentica di spiegarci come mai, troppo spesso, le persone vengono trattate come mezzi e non come fini, e chi sono gli attori che, storicamente, ci hanno ridotto a mezzi per i propri fini.

Ciò che Floridi considera “giusto” fare con l’IA, potrebbe essere impossibile da realizzare, se non cambia il sistema di sfruttamento dei dati su cui si basa il capitalismo digitale, o se non aumenta la competenza digitale dei cittadini, o se permangono le asimmetrie di potere tra grandi aziende tecnologiche e cittadini.

Il libro di Floridi rappresenta un tentativo approfondito, documentato e innovativo di stabilire dei confini etici per l’uso dell’IA e può sicuramente contribuire a un impiego più attento e meno rischioso di queste tecnologie, utile per i decisori politici (a cui principalmente si rivolge) e utile al dibattito pubblico, perché ci mostra che esistono anche notevoli benefici dall’applicazione dell’IA.

Però ha il limite di “dimenticarsi” del contesto socio-politico in cui finora l’IA è cresciuta.

(Una versione più estesa di questa recensione uscirà per la Rivista del Mulino)

Se continuiamo a tenere vivo questo spazio è grazie a te. Anche un solo euro per noi significa molto.
Torna presto a leggerci e SOSTIENI DOPPIOZERO

Raffaello Cortina Editore

Luciano Floridi

Etica dell'intelligenza artificiale

Sviluppi, opportunità, sfide



SCIENZA E IDEE

Collana fondata da Giulio Giorello