Ti con zero
scienza&arte, economia&politica&socie
recensioni di saggi, di letteratura, racconti,

PierLuigi Albini

229. Recensioni di saggi
Etica dell'intelligenza artificiale

Luciano Floridi

Etica dell'intelligenza artificiale
Sviluppo, opportunità, sfide

Raffello Cortina
2022
pp. 384

L'autore è uno dei non numerosi filosofi che si occupano delle tecnologie dell'informazione (ICT) sapendo però di cosa si parla, anche dal punto di vista tecnico. Non a caso è stato chiamato a far parte di importanti consessi internazionali per l'analisi delle opportunità e dei problemi di cui si tratta nel libro e ha contribuito alla stesura di importanti documenti ufficiali e di linee-guida di organismi sovranazionali. Questo è un libro - non userò l'abusato aggettivo di 'complesso' - che richiede un certo impegno di lettura. Potrebbe essere considerato come un vero e proprio manuale o come una miniera di informazioni sul tema dell'Intelligenza artificiale (IA); oppure come un testo filosofico, per l'appunto, ma anche come una tavola di suggerimenti e proposte per i decisori politici (per lo più latitanti, nella realtà italiana in particolare, vista la scarsissima attenzione dei programmi elettorali su tecnologie e digitale).[1] L'ampiezza del testo obbliga, quindi, a una recensione più lunga, anche se non potrà comunque essere esauriente, considerata la ricchezza del materiale offerto dal testo.

Floridi scrive che non affronta nel libro le questioni di politica, a proposito di IA, riservandosi di parlarne nel suo prossimo scritto. Ma già in questo testo, specialmente se lo si legge assieme al precedente *Il verde e il blu. Idee ingenue per migliorare la politica*,[2] ci sarebbe materiale per impostare una politica all'altezza di quella che si presenta ed è a tutti gli effetti come una radicale rivoluzione: quella dell'avvento e della diffusione dell'Intelligenza artificiale. Da questo punto di vista, il libro è anche un'agenda per le politiche energetiche e ambientali offerta ai decisori politici e ai privati.

Floridi espone chiaramente la sua posizione sulle prospettive future della IA, né con i catastrofisti né con le fughe fantascientifiche in avanti ma, come si dice nel sottotitolo, la ritiene una opportunità (se la si governa) e, allo stesso contempo, una sfida che, se non correttamente affrontata, potrebbe ge

[1] Dall'analisi comparativa di Wired sui programmi elettorali dei partiti sull'argomento: "Mancano idee concrete, i temi sono abbozzati e le parole usate in libertà. Le maggiori forze politiche non sembrano interessate a un tema sempre più centrale per lo sviluppo economico e sociale del Paese".
[2] Il verde è l'ambiente e il blu la tecnologia digitale e, in particolare, l'IA, che può fattivamente collaborare.

nerare guai. Scrive l'autore che "la discussione sul bicchiere mezzo vuoto e mezzo pieno [*a proposito dell'IA*] è inutile perché la questione davvero interessante è come possiamo riempirlo […] e il primo fondamentale passo è cercare di averne una maggiore e migliore comprensione". Nel riprendere poi un'affermazione di E.W Dijkstra, un noto informatico olandese, che "la questione se un computer possa pensare non è più interessante della questione se un sottomarino possa nuotare", l'autore si concentra sul merito delle questioni, parlando anche dettagliatamente della, per così dire, biografia della IA e del che fare.

Ora, la fonte prossima di tante incomprensioni e anche di inutili dibattiti sulla IA è proprio il termine di 'intelligenza'. Il fatto è che non esiste una nozione condivisa di intelligenza. Esistono solo alcuni indicatori, specialmente indagati dalle neuroscienze e dagli studi comparati. Se, per dirne alcuni, indichiamo come intelligente la comprensione semantica e quella delle cose, la preoccupazione per quelle stesse cose e per gli effetti delle nostre azioni, se è l'essere dotati di preferenze e di passioni per qualcosa (per non parlare di altri sentimenti e dell'autocoscienza), allora l'IA non è intelligente. Anzi, non è "più intelligente di un tostapane", scrive Floridi. Insomma, cosa è, con una definizione scientifica e onnicomprensiva, l''intelligenza'? alla domanda si può rispondere un po' sbrigativamente con un: boh! Alle neuroscienze e dintorni spetta la possibile risposta futura.

Però l'espressione 'intelligenza artificiale' ha preso ormai piede, ma secondo l'autore non è un termine scientifico "come 'triangolo', 'pianeta', o 'mammifero' […] è un'espressione generica, proprio come 'amicizia' o 'pornografia'". Perciò - e questa potrebbe essere una definizione più scientifica – sarebbe allora più corretto chiamarla, secondo me, come "operatività artificiale" (OA), se Floridi ha ragione. E ciò perché, in primo luogo, l'IA funziona in modo sostanzialmente diverso dall'intelligenza umana. Per esempio, "il principale cavallo di battaglia dell'IA non è più la deduzione logica ma l'inferenza e la correlazione statistica". Per mettere poi da parte la fondamentale differenza tra il biologico e l'hardware che supporta la IA.[3] Più avanti vedremo anche la questione della matematica degli algoritmi e dei sistemi di apprendimento (*machine learning*). Insomma, secondo l'autore "la vera IA non è logicamente impossibile, ma è assolutamente non plausibile".

A questo proposito, il punto chiave della esposizione di Floridi è nella interpretazione di cosa è la IA, piuttosto differente da tanti sociologismi e opinioni correnti sui 'social'. Si tratta di un concetto che nel libro viene ripetuto più volte in vari modi (perché ci entri bene in testa?), ma vale la pena di citarne alcuni. La IA è: la "separazione/scissione tra risoluzione di problemi e il comportamento intelligente"; è "come una nuova forma di disallineamento digitale tra azione e intelligenza, un fenomeno senza precedenti che ha causato parecchi malintesi"; si tratta di "una nuova forma di agire e non di intelligenza"; e ancora, "nell'IA, è il risultato che conta, non se l'agente o il suo comportamento sia intelligente". Per questo, "**l'IA non concerne la capacità di riprodurre l'intelligenza umana, ma in realtà la capacità di farne a meno**". [*il neretto è mio*] Infine, va tenuto in conto che nessuna IA, nelle sue varie versioni, ha mai superato davvero il test di Turing.[4]

In altra parte del testo, l'autore scrive che "l'IA non dovrebbe essere interpretata come un matrimonio tra una intelligenza di tipo biologico e artefatti ingegnerizzati, ma come un divorzio tra l'agire e l'intelligenza, cioè una scissione tra la capacità di affrontare problemi e compiti con successo in vista di uno scopo e l'esigenza di essere intelligenti". Qui, però e a mio avviso, matrimonio e: divorzio paradossalmente convivono, per cui le due interpretazioni potrebbero essere considerate complementari. In effetti, è lo stesso Floridi, nell'affrontare la questione dei rischi, a temere – richiamando Kant – che "gli esseri umani possono diventare inavvertitamente parte del meccanismo". Che siano gli esseri umani a doversi adattare a un nuovo ambiente, l'autore lo dice anche nel caso delle tanto decantate *smart cities*: "quando parliamo di città smart facciamo riferimento anche al fatto

---

[3] Qui, poi sarebbe necessario aprire una riflessione sul fatto che nel caso biologico ci troviamo di fronte a un sistema compatto, per cui non c'è distinzione tra hardware e software; insomma, si pensa con tutto il corpo, superando la tendenza al dualismo tipico della nostra cultura, che è invece praticabile nella distinzione tra hardware e software.

[4] Il test originario di Turing è stato riformulato nel corso degli anni, affinandone l'affidabilità.

che stiamo trasformando gli habitat sociali in luoghi in cui i robot possono operare con successo". Insomma, Floridi si chiede se "è il mondo che si sta adattando all'IA e non viceversa". Ma, allora, se di matrimonio si tratta, "dato che l'IA è il coniuge stupido ma laborioso e l'umanità quello intelligente ma pigro, chi si adatterà a chi?" Questa è una domanda chiave e anche inquietante, se l'IA non verrà governata, come si propone nel seguito del saggio. Insistendo su tale questione, l'autore, distingue le forme di intelligenza tra quella 'riproduttiva', ingegnerizzata, della IA e quella 'produttiva', biologica; e ribadisce che è "il mondo che si sta adattando all'IA e non viceversa" e che è "l'ambiente che è progettato in modo tale da essere compatibile per i robot, non il contrario". Beh, se questo non è un matrimonio…. Pensiamo, poi, alle protesi personali sempre più efficienti e dotate di IA. In questi casi, si potrebbe fare riferimento al movimento transumanista – anche se l'autore non ne parla, ritenendo forse che appartenga a uno dei filoni culturali delle non condivisibili "fughe fantascientifiche in avanti".

Ora, va precisato che l'adattamento umano a nuove tecnologie non è una novità e che è iniziato ben prima di *Homo sapiens* (con buona pace dei tecnofobi e dei seguaci fanatici del 'naturale') e che dobbiamo considerarlo come un dato di fatto dell'evoluzione culturale generata da quella naturale. Tuttavia, nel caso dell'IA ci troviamo di fronte a un evento completamente nuovo, molto più penetrante e invasivo rispetto a tutte le precedenti tecnologie. Qualcuno, a ragione, ha parlato di **rivoluzione antropologica**, il cui esito nessuno può conoscere. Si possono solo fare delle ipotesi, ma niente previsioni. Forse il massimo che si può fare è disegnare degli scenari possibili o probabili.

La questione degli algoritmi – sulla cui definizione c'è però un accordo limitato - rappresenta il cuore delle preoccupazioni di Floridi; basti pensare alla loro struttura e alla loro implementazione attraverso il cosiddetto *deep learning* che macina quantità incredibili di dati, completamente al di là delle capacità umane. Il pericolo di questa tecnologia consiste nell'essere fuori da ogni controllo e priva di una validazione etica. Tanto più che per i dati forniti alla IA per il *deep learning* la "qualità è importante, la provenienza è fondamentale" – osserva l'autore. Infatti, "è risaputo che gli algoritmi non sono eticamente neutri". Per dirla in breve, se i dati forniti non sono attendibili, se contengono delle distorsioni e dei pregiudizi, se sono concepiti per fini non leciti, possono fare e fanno parecchi danni. Per esempio, si possono "trasformare i dati in prove per un dato risultato oppure per innescare e motivare un'azione che può avere conseguenze etiche". Per esempio, quelli utilizzati da alcune polizie statunitensi contengono – come è stato accertato – pregiudizi razziali e di genere. Del resto, è proprio la necessità di definire un'etica condivisa il presupposto affinché questa tecnologia non vada fuori controllo. A tale proposito, il libro è molto circostanziato e articolato, come è ovvio fin dal titolo.

Floridi discute le diverse forme non-etiche e anche criminali in cui può essere impiegata l'IA, parecchie della quali sono già in circolazione nella infosfera. Di fronte al proliferare di proposte di codici etici, che rischiano di generare solo confusione e favorire un impiego scorretto dell'IA, l'autore rileva che basterebbe concordare sull'applicazione di quattro principi più uno: i primi quattro sono gli stessi adottati dalla bioetica (beneficienza, non maleficenza, autonomia, giustizia); il quinto, specifico per la sfera digitale, è l'esplicabilità ovvero, per estensione, la intellegibilità, la trasparenza e la responsabilità. Questo quinto principio è il cardine di molte proposte contenute nel libro a proposito degli algoritmi, di cui si sollecita "un sistema di certificazione per prodotti e servizi digitali". Si tratta di emulare ciò che si è finora ottenuto nell'etica ambientale.

In buona sostanza, Floridi è per quello che definisce un "design partecipativo", una sorta di contratto sociale, "un patto tra le diverse parti interessate umane, mediato dalle macchine". L'autore, inoltre, ci mette in guardia sul fatto che "l'IA non è un altro dispositivo che deve essere regolamentato una volta che sia stato sviluppato. Abbiamo commesso l'errore di pensare in questi termini quando il Web ha iniziato a crescere negli anni Novanta e oggi il pasticcio è evidente". Insomma, come nel caso del riscaldamento globale, **occorre intervenire ora** e siamo già in grave ritardo.

Ma di chi è la responsabilità se applicando la IA qualcosa va storto? Come si dice in gergo, si tratta di una tecnologia "dalle molte mani", perciò – secondo alcuni - sarebbe difficile o impossibile deciderne la responsabilità. Eppure, l'autore, dopo un'attenta disamina dei casi, approda a una conclusione condivisibile: "se il design è imperfetto e l'esito illecito, allora tutti gli agenti (umani) coinvolti debbono ritenersi responsabili". Programmatori, produttori e utilizzatori; ma anche committenti, se la committenza si prefiggeva fini illeciti. Naturalmente il testo si sofferma sul complesso dei problemi posti dagli usi per il male e da quelli per il bene, come, per citarne solo uno, la questione della *privacy*.

Le questioni sociali e ambientali ricevono un'attenzione particolare, avanzando anche proposte concrete. Per esempio, parlando degli impatti sociali, Floridi scrive che "se non elaboreremo migliori quadri normativi giuridici ed etici, l'IA creerà ulteriori polarizzazioni nella nostra società, specialmente tra i pochi al di sopra della macchine – i nuovi patrizi – e coloro che stanno al di sotto – la nuova plebe". Per questo - osserva – "occorrerà fare qualcosa al riguardo, facendo pagare più tasse alle aziende e ai benestanti". Immagino le contrarietà dei neoliberisti, i quali, naturalmente si guardano bene dall'entrare nel merito di quella che si è definita una vera e propria rivoluzione antropologica indotta dal digitale, come del resto sul che fare davvero per contrastare il cambiamento climatico.

Aggiungo che servirebbe uno sforzo coordinato maggiore nel campo educativo e della formazione digitale, anche per gli adulti, affinché i cittadini non siano utenti passivi delle tecnologie e possano sviluppare una capacità critica, ancorché limitata. Il discorso richiederebbe un più ampio sviluppo, anche se dal libro si possono ricavare anche delle indicazioni operative. Ma i problemi ci sono anche nel mondo del lavoro – e l'autore ne parla. Alcuni di tali problemi sono stati segnalati da Luigi Agostini e Marcello Malerba, nell'articolo pubblicato quattro anni fa su Ticonzero *Il lavoro prima e dopo l'algoritmo*.[5]

Non che non esistano progetti e piani di intervento, a livello nazionale e europeo. Ma, per esempio, nel Piano Nazionale Scuola Digitale non si parla mai degli specifici problemi relativi alla IA;[6] mentre nel Programma Strategico Italiano Intelligenza Artificiale non mi sembra che le questioni etiche ricevano la dovuta attenzione. Meglio sta facendo l'Europa adottando, dopo un lungo lavoro di elaborazione e di consultazioni, delle linee guida secondo la quali "La Commissione propone nuove norme per garantire che i sistemi di IA utilizzati nell'UE siano sicuri, trasparenti, etici, imparziali e sotto il controllo umano. Pertanto, essi sono classificati in base al rischio".

Si è accennato sopra alla rivoluzione antropologica che viene tuttora sottovalutata, per deficit di approccio degli studiosi di varie discipline e/o per disattenzione o incultura dei decisori politici. In realtà, "ciò che conta – per concludere con l'autore – "non sono tanto i bit in movimento invece degli atomi – questa è un'interpretazione della società dell'informazione obsoleta e basata sulla comunicazione che deve troppo alla sociologia dei mass media - quanto piuttosto il fatto molto più radicale che il nostro modo di comprendere e concepire l'essenza e il tessuto della realtà sta cambiando. In effetti, abbiamo cominciato ad accettare il virtuale come parzialmente reale e il reale come parzialmente virtuale".

1° settembre 2022

Codice ISSN 2420-8442

[5] Con la collaborazione di Gianni Marchetto e Duccio Campagnoli e la revisione di PierLuigi Albini

[6] Ci sono anche programmi e lezioni sull'informatica a partire dalla scuola primaria, come "L'ora del codice", ma non sembra che si impari a programmare (algoritmi) quanto piuttosto a usare gli strumenti esistenti