

"Etica dell'intelligenza artificiale" di Luciano Floridi

di Michele Veneziano

Recensione a: Luciano Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina Editore, Milano 2022, pp. 340, 26 euro (scheda libro)

Luciano Floridi è Professore ordinario di Filosofia ed Etica dell'informazione all'Università di Oxford, dove dirige il Digital Ethics Lab, e di Sociologia della cultura e della comunicazione all'Università di Bologna. È inoltre chairman del Data Ethics Group dell'Alan Turing Institute, l'istituto britannico per la data science. Nei numeri cartacei di Pandora Rivista abbiamo pubblicato due sue interviste dal titolo "Filosofia dell'infosfera" e "Piattaforme digitali: una prospettiva filosofica"; abbiamo inoltre avuto il piacere di ospitarlo per un appuntamento online dei Dialoghi di Pandora Rivista e alla prima puntata del nuovo videopodcast dei Dialoghi di Pandora Rivista intitolata "Governare l'intelligenza artificiale" e dedicata ai problemi concettuali posti dallo sviluppo del digitale e dell'IA. Proponiamo qui la recensione del suo ultimo libro, uscito a marzo 2022 per Raffaello Cortina Editore.

Il nuovo libro di Luciano Floridi - *Etica dell'intelligenza artificiale* - rappresenta un ulteriore tassello nella tetralogia sui fondamenti della filosofia dell'informazione (*Principia Philosophiae Informationis*), impresa che l'autore porta avanti da diversi anni e che ha contribuito a renderlo uno dei filosofi più apprezzati a livello mondiale. Dopo una breve introduzione, il presente contributo traccia una panoramica dei contenuti del volume e, in conclusione, offre alcune osservazioni sul testo.

Originariamente, l'obiettivo del libro - racconta Floridi nella prefazione - doveva essere quello di indagare le nuove forme dell'agire artificiale e politico indotte dalla rivoluzione digitale. L'autore ha però deciso di suddividere questo arduo compito in due volumi: l'opera in questione, che si sofferma sull'agire artificiale, e un secondo volume che vedrà la luce nei prossimi anni e si focalizzerà sulla politica dell'informazione e l'agire politico, inteso come forma di agire collettivo influenzata dalle interazioni digitali.

Etica dell'intelligenza artificiale offre una estesa trattazione concettuale dell'intelligenza artificiale (IA), considerata pragmaticamente come una normale tecnologia digitale. Questa tecnologia è però caratterizzata da una importante peculiarità, ovvero l'essere una nuova forma di agire efficace «ma non intelligente» (p. 65), resa possibile dalla rivoluzione digitale o, più precisamente, dal «disallineamento digitale tra azione e intelligenza» (p. 21). Come spesso accade quando ci si avvicina a testi che trattano il tema dell'IA, anche in questo caso alla fine il lettore si ritrova inevitabilmente a domandarsi se il termine "intelligenza artificiale" sia qualcosa di più in un seducente termine-ombrello che nella pratica si rivela piuttosto inadatto per descrivere il reale funzionamento dei sistemi digitali a cui fa riferimento. La tesi al cuore del volume appare coerente con questa osservazione: secondo Floridi, infatti, l'IA rappresenta un divorzio tra intelligenza e capacità di agire, reso possibile principalmente da due fenomeni. Da una parte, l'avanzamento delle tecnologie digitali che permettono di separare la capacità di risolvere un problema con successo dall'esigenza di essere intelligenti nel farlo. D'altra parte, la crescente tendenza da parte della società a trasformare il mondo in un ambiente sempre più adatto al funzionamento dell'IA - fenomeno che l'autore definisce "avvolgimento" - contribuisce altresì all'affermazione di questi sistemi tecnologici.

Dovrebbe già essere chiaro al lettore, dopo queste poche righe, che quella che Floridi fornisce non

è dunque semplicemente un'introduzione all'IA o all'etica dell'IA, ma una vera e propria trattazione concettuale e interpretativa del significato profondo e delle conseguenze sociali di queste tecnologie. Questo volume non è, dunque, il testo adatto per il lettore in cerca una introduzione all'IA. Probabilmente questo libro non è nemmeno il luogo adatto per chi cerca un'introduzione al pensiero di Floridi, soprattutto perché il ruolo dell'informazione e i concetti fondamentali da lui proposti nel corso degli anni (come le nozioni di onlife, inforg e infosfera o una trattazione esaustiva della rivoluzione digitale) sono appena accennati, nonostante siano in realtà fondamentali per comprendere come si sia arrivati all'IA e il contesto in cui questa tecnologia di inserisce[1].

Un'ultima nota da tenere in considerazione prima di passare ad una panoramica sul contenuto dei singoli capitoli è che l'IA, per Floridi, non rappresenta qualcosa da temere o ostacolare, ma piuttosto una "semplice" tecnologia, che necessita di lucidità ed etica per essere modellata, direzionata e governata - in modo che sia utilizzata per produrre e sostenere il bene sociale -. È proprio in quest'ottica che la filosofia e l'etica devono essere utilizzate per fare ciò che l'autore definisce "design concettuale", ossia come strumenti a sostegno della governance delle tecnologie, che devono indicare la strada eticamente più sostenibile e più giusta.

La prima parte del libro (intitolata *Comprendere l'intelligenza artificiale*) è composta da tre capitoli su passato, presente e futuro dell'IA, che contengono le fondamenta concettuali su cui si costruisce il volume. In questi capitoli, viene indagata l'origine dell'IA da un punto di vista concettuale, soffermandosi su come la rivoluzione digitale abbia creato le condizioni per lo sviluppo di questi sistemi grazie al suo "potere di scissione", ovvero la capacità delle tecnologie digitali di scindere e ricomporre realtà e idee (secondo una sorta di logica del "taglia e incolla"), che porta a una re-ontologizzazione e ri-epistemologizzazione della modernità. Queste ultime non sono certamente idee nuove nel pensiero di Floridi e sono state ampiamente elaborate nei suoi lavori precedenti. È in questo quadro che l'autore definisce l'IA come una nuova forma di agire intelligente (o più precisamente, come riserva di capacità di agire) determinata dal disallineamento digitale tra azione e intelligenza, e spiega come il design concettuale possa essere la controparte del potere di scissione del digitale, in un costante gioco di pesi e contrappesi. L'interpretazione dell'IA di Floridi è squisitamente filosofica e si contraddistingue rispetto alle interpretazioni offerte dalle tradizioni ingegneristiche e cognitive, che hanno affrontato largamente il tema fino dalla seconda metà del secolo scorso. In questi primi capitoli viene introdotto inoltre il concetto di "avvolgimento", spiegando come il successo dell'IA sia in gran parte dovuto al fatto che la società sta gradualmente costruendo un ambiente sempre più adattato ad essa. Il rischio che emerge è evidentemente quello che l'umanità si adatti alla tecnologia e non viceversa. Floridi si interroga poi sui possibili sviluppi dell'IA e sostiene che questi dipenderanno in larga parte da due elementi. In primo luogo, la capacità di utilizzare piccoli dati di alta qualità, assieme alla crescente capacità dell'IA di generare i propri dati sintetici (cioè auto-generati sulla base delle regole predefinite) in particolare in quei contesti che è possibile "ludicizzare", ovvero formalizzare come giochi. In secondo luogo, dalla capacità di tradurre i problemi in modo che siano risolvibili per l'IA, che significa fondamentalmente trasformare problemi difficili in problemi complessi da un punto di vista computazionale[2], ricordando che non ha senso «[...] cercare di imitare gli esseri umani attraverso l'IA. Dobbiamo sfruttare, invece, ciò che le macchine, inclusa l'IA, fanno meglio. La difficoltà è nemica delle macchine, la complessità il loro alleato» (p. 83).

La seconda parte del libro è invece intitolata *Valutare l'intelligenza artificiale* e si articola nei restanti undici capitoli. L'obiettivo generale è quello di discutere le conseguenze delle concettualizzazioni e argomentazioni portate nella prima parte sotto diversi punti di vista. Se la prima parte del volume ha un'impostazione metateorica e interpretativa, la seconda risulta più analitica. Questi capitoli, anche se posizionati seguendo una logica ben evidente, sono sufficientemente indipendenti da permettere di essere affrontati singolarmente, senza la necessità di seguire l'ordine prestabilito. I capitoli

quattro, cinque e sei sono incentrati sui principi etici dell'IA e offrono rispettivamente: i risultati di un'analisi comparativa di alcuni degli insiemi di principi etici auspicati da documenti di importanti organizzazioni internazionali, che permette all'autore di elaborare una prospettiva unificata sui principi etici dell'IA (capitolo 4); una discussione sui rischi che derivano dalla non corretta applicazione di questi principi come lo shopping etico, il bluewashing etico, il lobbismo etico, il dumping etico e l'elusione dell'etica (capitolo 5); e un'analisi della relazione tra governance, etica e regolazione digitale, che porta ad auspicare l'utilizzo di una etica soft - intesa come etica post-compliance - nell'implementazione di sistemi di IA (capitolo 6). Nel capitolo 7 vengono analizzate le sfide etiche poste dallo sviluppo di IA attraverso la mappatura etica degli algoritmi, strumento che mette alla luce le questioni epistemiche (prove inconcludenti, prove imperscrutabili, prove fuorvianti), normative (esiti ingiusti, effetti trasformativi) o relative alla tracciabilità. Il capitolo 8 offre un'analisi sistematica dei crimini di IA e sottolinea come la responsabilità di fondo di questi nuovi crimini sia degli esseri umani che sfruttano l'IA in modo contrario all'etica e al diritto. Nel capitolo 9, dedicato alle buone pratiche nell'applicazione dell'IA, si definisce il concetto di IA per il bene sociale (AI4SG) e si cerca di mostrare le caratteristiche eticamente rilevanti dei progetti che potrebbero essere descritti in termini di AI4SG (falsificabilità e implementazione incrementale; garanzie contro la manipolazione dei predittori; intervento contestualizzato in ragione del destinatario; spiegazione contestualizzata in ragione del destinatario e finalità trasparenti; tutela della privacy e consenso dell'interessato; equità concreta; e semantizzazione adatta all'umano). Nel capitolo 10 viene criticata la tendenza a parlare di IA come singolarità tecnologica e viene rimarcata l'importanza della lungimiranza etica e di design per modellare una società più giusta. Nel capitolo 11 sono poi esaminate più nel dettaglio alcune raccomandazioni per valutare, sviluppare, incentivare e sostenere una AI4SG. Il capitolo 12 discute gli impatti positivi e negativi dell'IA sull'ambiente il potenziale di queste tecnologie per contrastare i cambiamenti climatici. Nel capitolo 13 vengono poi approfondite le possibilità di avvalersi di IA per sostenere gli obiettivi di sviluppo sostenibile delle Nazioni Unite. L'ultimo capitolo è infine dedicato alla necessità di un nuovo matrimonio tra il verde (ovvero le questioni ambientali) e il blu (ovvero le questioni tecnologiche), in cui Floridi porta le sue conclusioni e introduce il futuro volume sulla politica dell'informazione.

Il testo di Floridi si caratterizza per eleganza e chiarezza espositive, elementi d'altronde trasversali nella sua vasta produzione saggistica. Nonostante la complessità di alcune questioni a livello concettuale, risulta evidente lo sforzo per rendere il libro il più possibile avvicinabile anche ai non esperti. Floridi intende, prima di ogni altra cosa, essere capito e non nascondere le sue idee dietro complesse strutture grammaticali o tecnicismi filosofici. Coerentemente con il suo obiettivo di sviluppare una filosofia che possa contribuire a migliorare il mondo e la politica, Floridi continua con la sua impresa di design concettuale e a volte pare essere più interessato a parlare a chi realmente si occupa di IA (politici e settore privato in primis) piuttosto che all'accademia. L'operazione è tendenzialmente riuscita, in particolare nella prima parte del volume. Il lettore sia consapevole che il libro rimane comunque un testo impegnativo, soprattutto nelle sezioni più dense di filosofia del diritto. La chiarezza con cui Floridi porta le sue argomentazioni non equivale naturalmente al dover essere in accordo con lui in tutte le sue elaborazioni teoriche, ma rimane un pregio, perché offre la possibilità al lettore di essere in disaccordo, rendendo la lettura intellettualmente vivace.

Floridi dimostra in questo libro che aprire la "scatola nera" dell'IA è un'impresa tutt'altro che semplice. Non solo perché all'interno vi si troveranno complesse operazioni computazionali, ma anche (e forse soprattutto) perché l'IA obbliga la società a interrogarsi su cosa sia eticamente giusto. Floridi, inoltre, non si limita semplicemente ad aprire questa scatola nera, ma ci ricorda che nonostante non sia sempre semplice comprendere del tutto il funzionamento di un sistema di IA, ciò non ci solleva dalla responsabilità di gestire questa tecnologia nel migliore dei modi possibili. Per farlo abbiamo a disposizione lo strumento del design, contrappeso fondamentale per gestire

l'impatto dell'IA sulla società - ovvero l'arte di risolvere un problema sfruttando vincoli e possibilità, per soddisfare alcuni requisiti in vista di un obiettivo. Il design ha però bisogno di un progetto comune e condiviso, e l'etica è lo strumento che può aiutare a distinguere il buon design dal cattivo design. Nella battaglia quotidiana contro il determinismo e il soluzionismo tecnologico, che continuano ad alimentare il discorso di molti promotori dell'IA sia pubblici che privati, Floridi ricorda che siamo noi - ovvero la società, attraverso lo strumento della politica - gli unici artefici del nostro futuro. Sta a noi, dunque, l'impegno di trovare la lucidità necessaria per governare l'IA. Etica dell'intelligenza artificiale di Luciano Floridi offre una interpretazione utile per farlo.

[1] Per chi cerca un'introduzione al pensiero di Floridi, testi più adatti sono sicuramente: La quarta rivoluzione. Come l'infosfera sta trasformando il mondo, edito da Raffaello Cortina Editore nel 2017 e La rivoluzione dell'informazione (traduzione di Information: A Very Short Introduction) del 2010, edito in Italia, da Codice Edizioni nel 2012, con una prefazione di Juan Carlos De Martin.

[2] Il termine "complesso" utilizzato da Floridi fa riferimento alla teoria della complessità computazionale, una branca dell'informatica che si concentra sulla classificazione dei problemi computazionali in base al loro utilizzo delle risorse (ad esempio, il tempo di calcolo e la quantità di memoria richiesta per svolgere un compito).