

"Intelligenza Artificiale" di Luciano Floridi e Federico Cabitza

di Otello Palmi

Recensione a: Luciano Floridi e Federico Cabitza, *Intelligenza Artificiale. L'uso delle nuove macchine*, Bompiani, Milano 2021, pp. 192, 12 euro (scheda libro)

«Non rimprovero all'intellettuale americano il suo atteggiamento ostile verso l'età della scienza e della macchina. Un atteggiamento ostile è qualcosa di produttivo e di positivo, e vi sono molti aspetti nelle grandi trasformazioni suscitate dall'età della macchina che hanno bisogno di una resistenza attiva e intelligente. Lo rimprovero per la sua mancanza di interesse per l'età della macchina. Egli non la considera così importante da indurlo a conoscere i fatti principali della scienza e della macchina e a prendere una posizione attiva nei confronti di essi» (Norbert Wiener, *Introduzione alla Cibernetica* [1]). Sostituendo «americano» con "italiano" e consapevoli di quanto il "mondo della macchina" sia cambiato negli ultimi settant'anni, tanto è passato da *The Human Use of Human Beings* di Wiener, possiamo comprendere la logica e l'importanza nel nostro dibattito culturale di questo breve libro che raccoglie gli interventi di Federico Cabitza e Luciano Floridi sul tema dell'Intelligenza Artificiale e del suo uso. Questi interventi si sono svolti il 14 ottobre 2021 all'annuale Martini Lecture Bicocca che prosegue la volontà del cardinale Carlo Maria Martini di dialogo con le avanguardie del pensiero scientifico e di discuterne l'impatto sulla società e sulla coscienza degli uomini.

Il ragionamento di Cabitza, Professore associato di Interazione Uomo-Macchina all'Università di Milano-Bicocca, inizia riconducendo la tecnologia all'interno dell'orizzonte del fare umano. La prassi umana intrattiene con il contesto tecnologico un rapporto articolato: infatti ne è in qualche modo la sorgente - la tecnologia resta una prassi umana - ma ne risulta a sua volta influenzata, infatti il contesto tecnologico condiziona evidentemente le possibilità del fare umano. Per comprendere in maniera propria il contesto tecnologico in cui viviamo è utile iniziare con il riferirsi all'orizzonte della macchina come «semplicemente qualsiasi mezzo, artificio o meccanismo (comprese regole, protocolli, gerarchie e divisioni del lavoro) che rende tutto questo [varie attività della prassi umana] possibile anche quando non lo sarebbe per come è fatto il nostro corpo» (p. 22). La macchina, dunque, è ciò che si inserisce all'interno del processo dell'agire umano integrandolo e conferendogli una capacità che le sue facoltà fisiche - o forse sarebbe meglio dire psico-fisiche - non rendevano possibile.

Tuttavia, quando parliamo di Intelligenza Artificiale (AI) ci riferiamo ad una precisa declinazione storico-concettuale-tecnica di questo ampio concetto di macchina. Siamo nel contesto delle tecnologie della computazione che «operano e muovono effettivamente cose immateriali, rappresentazioni di segni e quindi idee; ma ciò che fanno rende possibile l'estrazione, la lavorazione e il movimento di materia, ci permette di contare, ordinare, disporre degli oggetti, degli uomini e le loro azioni» (pp. 24-25). Queste tecnologie, insomma, sono iscritte nel mondo materiale e ne innervano ormai le relazioni: tanto quelle tra umani quanto quelle tra gli umani e le cose. In definitiva, nota Cabitza «La AI riguarda attività - automatizzabili in parte o interamente - di scelta, tra un numero limitato di opzioni, categorie, parole, valori; azioni che sono volte a raggiungere obiettivi complessi in contesti reali caratterizzati da un alto contenuto di conoscenza, autonomia di giudizio e di incertezza» (pp. 33-34). Come notava già Wiener, uno dei padri del pensiero cibernetico, se la prima rivoluzione industriale implicava una sostituzione, o meglio una sostituibilità della forza fisica umana, in quella che già negli anni Cinquanta veniva presentata come la seconda rivoluzione industriale ciò che viene sottoposto ad un certo grado di sostituibilità è la capacità umana di

giudizio.

In questa prospettiva di profonda riorganizzazione dell'agire umano Cabitza, evitando un dibattito tanto articolato quanto talvolta sterile sulla questione della singolarità, evidenzia come non si debba mai dimenticare che, anche in questo nuovo contesto tecnologico, «è sempre l'uomo che organizza le cose per trarne un vantaggio» (p. 46). Infatti, attraverso i richiami fondamentali a Langdon Winner (*Do Artifacts Have Politics?* [2]) e Bruno Latour (*On Technical Mediation* [3]) l'accento è posto sul fatto che tutti gli artefatti e quindi anche quella particolare tipologia di macchine che chiamiamo Intelligenza Artificiale siano politicamente connotati, ovvero siano progettati, esprimano, aiutino a sviluppare alcuni valori, alcune interpretazioni del mondo. Questa essenza non neutrale dell'elemento tecnologico è un passaggio apparentemente ovvio ma decisivo in moltissimi dei recenti dibattiti sul ruolo epistemologico, politico e sociale dell'AI. La comprensione della logica di composizione interna di queste macchine e del modo in cui agiscono diventa decisiva per una corretta impostazione delle modalità di relazione ad esse. L'indagine su ciò che possiamo definire algoritmo diventa necessaria dato che «la sua esecuzione [dell'algoritmo] (o interpretazione) da parte di una macchina computazionale fa emergere un programma e questo esprime un funzionamento, cioè quello che alla luce di "determinate serie di obiettivi definite dall'uomo" è una funzione della macchina e, nel caso di macchine interattive (come sono molti sistemi di AI), dei comportamenti, attuati dall'ibrido uomo macchina, uomo-e-tecnologia-in-uso» (p. 49).

Lo studio e l'approfondimento di queste questioni diventa una necessità in una società in cui le AI hanno ormai un peso estremamente rilevante. A questa mediazione algoritmica con la realtà dobbiamo essere in grado di affiancare una riflessione sugli aspetti androritmici della nostra esperienza, ovvero su tutto ciò che non riteniamo analizzabile e risolvibile algoritmicamente [4]. La strutturazione della relazione possibile tra questi due ritmi - algo e andro - diventa allora una questione centrale per evitare che «nel lungo periodo le macchine AI, concepite inizialmente per potenziare le capacità peculiari degli uomini "a beneficio degli uomini" e, tra tutte, le capacità di giudizio e di interpretazione di certi casi complessi finiscano paradossalmente per ottenere un effetto contrario e opposto di "depotenziamento"» (p. 80). Per scongiurare questo rischio, che purtroppo in parte è già una realtà, è necessario agire nella sfera del design - termine centrale anche nella riflessione di Luciano Floridi - tanto delle macchine stesse (approccio *ethical by design*) quanto sulla cultura e sulla formazione di chi esegue le pratiche di progettazione di queste macchine (*ethical in design*). Cabitza fa riferimento al paradigma del principio di responsabilità strutturato da Hans Jonas per esprimere il nuovo quadro etico e le sfide che esso porta con sé nell'era dell'Intelligenza Artificiale. Dobbiamo essere responsabili e agire cercando di calcolare le conseguenze, spesso non evidenti ed esplicite, dell'azione di questa nuova tecnologia arrivando anche ad un approccio etico di *undesign* che permetta di inibire alcune funzionalità rischiose di certe tecnologie, di ritagliare degli spazi non tecnologicamente mediati, di prevedere di eliminare certe tecnologie da determinati ambienti in cui il loro apporto, seppur possibile tecnicamente, potrebbe risultare dannoso e precludendo lo sviluppo e l'utilizzo di specifiche tecnologie in determinati contesti (il riferimento è qui alle armi autonome).

Questi riferimenti e queste direzioni di ricerca ci sembrano preziose, come decisivo ci appare il richiamo al pensiero di Jonas. Infatti, come notava il pensatore tedesco, la nostra capacità tecnologica di azione sul reale ha finito per renderci responsabili anche di ciò che è fuori di noi ma che è in nostro potere (la natura ad esempio); tuttavia la specificità della macchina AI dovrebbe forse farci riflettere anche sul fatto che stiamo diventando responsabili della definizione stessa di cosa è umano e cosa non lo è, di quali azioni e modi di pensare sono tali e quali no in una maniera nuova rispetto al passato: nel bilanciamento tra algoritmo e androritmo è implicata una difficoltosa riflessione sulle capacità e sulle peculiarità umane, una riflessione di cui la comunità scientifica tutta deve sentire l'urgenza se è vero che, come sottolinea Cabitza, «una recente rassegna bibliografica

(BIRHANE et Al., 2021) ha notato che tra i cento articoli più citati negli ultimi quindici anni pubblicati negli atti delle due conferenze annuali più importanti del settore [dell'Intelligenza Artificiale] (NeurIPS e ICML), solo due lavori hanno menzionato e discusso, in qualche maniera, i potenziali danni derivanti dalle applicazioni presentate; gli altri novantotto lavori non ne hanno citato alcuno, così rimuovendo dal discorso la possibilità di discuterne [...]» (p. 74).

Per comprendere la logica in cui Luciano Floridi, Professore ordinario di Etica dell'Informazione all'Università di Oxford e di Sociologia della Comunicazione all'Università di Bologna, applica la sfera dell'etica alla nostra relazione con l'Intelligenza Artificiale è bene partire dal fatto che per il filosofo «L'etica non valuta solo per constatare, e magari premiare o punire il bene o il male, ma soprattutto per progettare: fare di più, fare meglio, magari non fare gli stessi errori» (p. 123). Questa caratterizzazione dell'etica come qualcosa di inerente alla progettazione è facilmente contestualizzabile all'interno dell'intervento che qui prendiamo in esame. Infatti Floridi divide preliminarmente il campo scientifico (una divisione funzionale al discorso in questo caso ma che lo stesso Floridi approfondisce in altri contesti) tra discipline mimetiche e poietiche: mentre le prime cercano di descrivere e rappresentare sistemi che scoprono nella realtà le seconde «non descrivono soltanto i loro sistemi, ma soprattutto li "disegnano", li "creano", o li "trasformano" cercando di migliorarli» (p. 134). Se tra le prime possiamo annoverare fisica, chimica e biologia tra le seconde ricordiamo ad esempio l'architettura e l'economia. Se le prime, semplificando, cercano di descrivere-rappresentare la realtà ontologica le seconde contribuiscono al suo accrescimento e alla sua modificazione. Questa distinzione è particolarmente importante per indirizzare la lettura di Floridi riguardo alla rivoluzione digitale in generale e all'AI in particolare, infatti «l'informatica, o computer science, come qui sarebbe meglio chiamarla, non "descrive" il mondo - anche se ormai nessuna scienza mimetica può fare a meno del suo supporto - ma lo "inscrive con simboli matematici": la computer science opera ontologicamente scrivendo nuove pagine, in codice binario, del libro dell'universo» (p. 137). Questa attività produttiva è ciò che guida l'interpretazione del digitale da parte di Floridi: la rivoluzione viene interpretata in questo contesto come un rivolgimento ontologico e non meramente epistemologico o addirittura comunicativo. È la struttura stessa del reale ad essere sottoposta ad uno sconvolgimento e non solo la nostra capacità di comprenderlo o di comunicare al suo interno.

Questa premessa è decisiva per l'interpretazione dell'AI. Anche Floridi, come Cabitza, mette da parte i dibattiti sulle super intelligenze o sulla singolarità per concentrarsi sulle forme di AI che stanno velocemente producendo effetti sulla nostra esperienza, ovvero l'Intelligenza Artificiale riproduttiva «che cerca con mezzi non biologici di ottenere l'esito (chiamiamolo output) del nostro comportamento intelligente, cioè risolvere problemi o svolgere compiti con successo in vista di un fine» (p. 139). Semplificando, possiamo dire che mentre gli sforzi per produrre un doppio, magari anche migliore, dell'intelligenza umana in formato digitale non sono andati troppo lontano, l'idea di far risolvere dei compiti, che se affrontati da esseri umani necessiterebbero di intelligenza, a delle macchine che non si pongono minimamente il problema di esserlo ha prodotto risultati strabilianti. Questo è stato possibile proprio perché il digitale non è semplicemente una rivoluzione epistemica ma anche e soprattutto ontologica: infatti non siamo riusciti ad inserire degli agenti artificiali intelligenti nel nostro ambiente con ottimi risultati in termini di problem solving, ma stiamo ristrutturando digitalmente l'ambiente per far sì che sistemi di problem solving profondamente diversi da noi abbiano successo. Nell'ontologia dell'infosfera (Floridi, La quarta rivoluzione [5]) ovvero in una realtà strutturalmente modificata da e in funzione del digitale le AI seppur non dotate di particolare intelligenza in senso umano (che resta l'unico senso laicamente noto) sono estremamente funzionali.

Iniziamo allora a comprendere per quale motivo l'etica applicata alle questioni digitali implica in primo luogo problemi di design e progettazione «Oggi giorno avvolgere l'ambiente in un'infosfera

compatibile con l'AI ha iniziato a pervadere tutti gli aspetti della realtà e sta accadendo quotidianamente ovunque, in casa, in piazza e in ufficio. Io stesso ho modificato il giardino per assicurarmi che il robot potesse tagliare l'erba senza incastrarsi in un angolo o uscire dal perimetro o bloccarsi in un'area scoscesa. Allo stesso modo quando parliamo di città intelligenti (smart cities) intendiamo anche dire che stiamo trasformando gli habitat sociali in luoghi in cui l'AI può operare con successo» (p. 157). Il concetto chiave diventa quello di envelope ovvero di "involucro": dobbiamo comprendere che le grandi capacità di azione di queste macchine riposano sul fatto che stiamo modellando il mondo in base alla loro capacità interpretative, stiamo riscrivendo ampi spazi di realtà in un linguaggio che permetta il loro funzionamento, creando degli involucri che ne massimizzino la resa.

Floridi tratta la questione nei termini di un divorzio tra intellighere e agere, un divorzio che non sarà retoricamente sensuale come la narrazione sulla singolarità, ma che sta permettendo di appaltare moltissime azioni, che fino a poco tempo fa richiedevano intelligenza e tempo umano per essere compiute con successo, a macchine che possono svolgerle, posta la re-ontologizzazione a cui abbiamo accennato, in modo molto più efficiente di noi. Questa nuova articolazione dell'agire nel contesto dell'infosfera porta con sé un rischio fondamentale che Floridi espone facendo riferimento ancora a Wiener. Il rischio è che «gli esseri umani possano diventare inavvertitamente parte del meccanismo comportando non solo una separazione tra agere ed intellighere, ma un'inversione della relazione tra i due, mettendo l'intellighere al servizio dell'agere» (p. 166). In primo luogo, sempre più spesso, questi sistemi artificiali hanno necessità di una componente interpretativa prettamente umana per superare alcuni problemi, tuttavia in questo caso le capacità umane vengono integrate in una logica artificiale dando luogo a quella inversione di priorità richiamata da Floridi «Questa tendenza abbastanza recente è nota come human-based computation [...] "un paradigma per utilizzare la capacità umana nel processare dati per risolvere problemi non ancora risolvibili dai computer" (ANH 2015)» (p. 167). In secondo luogo, gli umani sono invischiati in questo sistema anche come «interfacce influenzabili» ovvero sono esposti a servizi apparentemente gratuiti il cui vero obiettivo «è quello di fornire quello che serve per ottenere tutte le informazioni sull'interfaccia-cliente necessarie per garantire il grado di manipolazione che fornisce al business un accesso meno limitato e vincolato possibile al conto in banca» (p. 171). In questo senso ampio possiamo certamente dire, con Floridi, che il problema fondamentale è un problema di design, ovvero di strutturazione e articolazione degli spazi fisici e logici di interazione tra umanità e artificialità, tra intellighere ed agere, in questo spazio deve ritrovare centralità il pensiero filosofico, «dopotutto è un segno di intelligenza far funzionare la stupidità a proprio vantaggio. Intellighere deinde agere è meglio di agere sine intellighere» (p. 177).

Dai due interventi, insomma, emerge l'urgenza di una riflessione etica, politica e sociale sulla genesi e sull'uso di queste macchine. Una riflessione in grado di contestualizzarle nell'orizzonte economico-sociale-tecnico-filosofico in cui sono nate e in cui si stanno sviluppando. La comprensione della natura ontologica della rivoluzione digitale e la capacità di vedere uomini, e non macchine, dietro ai fini che l'agere delle AI persegue ci appaiono come condizioni necessarie per impostare correttamente questo importante orizzonte di dibattito. Il rischio è che l'enorme successo di questo agere artificiale possa portarci a pensare che l'intellighere sia in fondo superfluo o addirittura problematico in un orizzonte di pura efficienza in cui gli scopi, i perché e le possibilità di rivoluzionare i nostri modelli interpretativi divengano fattori apparentemente secondari. Seguendo Wiener «Allorché le persone umane sono organizzate come un sistema che li impiega non secondo le loro piene facoltà di esseri umani responsabili, ma come altrettanti ingranaggi, leve e connessioni, non ha molta importanza che la loro materia prima sia costituita da carne e da sangue. Ciò che viene usato come un elemento di una macchina, è un elemento nella macchina» (Introduzione alla Cibernetica [6]). Seguendo Jonas «In base a quello che affermano i cibernetici, la società è una

struttura comunicativa per la trasmissione, lo scambio e l'accumulo di informazioni e sono queste che di fatto la tengono insieme. Mai è stata proposta una nozione di società più vuota di questa. Non si dice nulla sul tipo di informazioni, e perché dovrebbe essere così importante averle. Lo schema proposto non lascia spazio neppure all'emergere di tali domande. Una qualsiasi teoria della socievolezza dell'uomo, per quanto cruda o distorta, che prenda in considerazione il suo essere una creatura di bisogno e di desiderio e che ricerchi i rapporti vitali che legano gli uomini, va molto a fondo della questione» (La Cibernetica e lo scopo: una critica) [7]. Crediamo che i riferimenti a Wiener e Jonas siano preziosi. Infatti, se il primo può essere interpretato come la volontà di portare avanti questa rivoluzione tecnologica così radicale con una grande attenzione alle sue ricadute sociali, il secondo può fornirci quell'atteggiamento filosofico di scepsi costruttiva (ci si perdoni la contraddizione) di cui il pensiero sulla tecnologia ha un disperato bisogno.

[1] Norbert Wiener, *Introduzione alla Cibernetica. L'uso umano degli esseri umani*, Bollati Boringhieri, Torino 2012, p. 176. Prima edizione: *The Human Use of Human Beings*, Houghton Mifflin, Boston 1950.

[2] Langdon Winner, *Do Artifacts Have Politics?*, in «*Daedalus*», Vol. 109, No. 1 (1980). Ristampato in: Donald A. MacKenzie e Judy Wajcman (a cura di), *The Social Shaping of Technology*, Open University Press, Londra 1985 e 1999; e anche riadattato in: Langdon Winner, *The Whale and the Reactor: A Search for Limits in an Age of High Technology*, University of Chicago Press, Chicago 1986.

[3] Bruno Latour, *On Technical Mediation*, in «*Common Knowledge*», Vol.3, n°2 (1994), pp. 29-64.

[4] Gerd Leonhard, *Tecnologia vs umanità. Lo scontro prossimo venturo*, Egea, Milano 2019.

[5] Luciano Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Raffaello Cortina Editore, Milano 2017.

[6] Norbert Wiener, *Introduzione alla Cibernetica. L'uso umano degli esseri umani*, op. cit., pp. 228-229.

[7] Hans Jonas, *La cibernetica e lo scopo: una critica*, Edizioni ETS, Pisa 1999, p. 60.