Che cosa ci spaventa nell'intelligenza artificiale?

Paolo Costa*

WHAT IS SO FRIGHTFUL ABOUT AI?

ABSTRACT: This is a decidedly Tocquevillian statement in which I give voice to my concern that the spread of AI may encourage a view of the (epistemic, moral, artistic) creativity of human beings as a refined, but nonetheless mechanical, variety of imitation. In my paper, I take seriously the widespread fears about the development of increasingly sophisticated AI devices in order to understand whether they convey some morally relevant intuitions. An unbridled use of algorithms would not be a big problem as such, if it did not take place within a society in which a form of disenchantment is systematically endorsed, that actually amounts to a tacit, but, precisely for this reason, even more disturbing debunking of humanity and its ability to change for the better.

KEYWORDS: Artificial intelligence; moral experience; moral justification; personal change; fears of science

1. Consulenza etica ed esperienza morale

In questo testo, scritto secondo lo stile assertivo dello *statement* scientifico, intendo occuparmi dei risvolti morali dei progressi nell'ambito dell'Intelligenza Artificiale. Questi ultimi includono anche quegli sviluppi tecnologici che sfumano il confine tra l'artificiale e l'organico, tra la macchina e l'organismo vivente, lasciando presagire scenari futuri in cui l'esercizio personale dell'intelligenza sarà reso possibile da organi, o parti di organi, che non sono più soltanto il prodotto sagace, ma non intenzionale, dell'evoluzione naturale.

Ora, la divisione del lavoro in ambito accademico prevede di regola che chi si occupa degli sviluppi teorici e pratici di tecnologie che hanno un impatto diretto sulla vita delle persone deleghi ad altri specialisti la loro giustificazione o confutazione morale. La giustificazione di credenze o comportamenti, tuttavia, non esaurisce il senso dell'esperienza morale umana. Al fianco di questo *côté* attivistico che punta a garantire agli agenti morali il pieno controllo delle premesse e conseguenze delle proprie scelte, c'è infatti nell'esperienza morale umana una dimensione ricettiva che, più che con la giustificazione, ha a che fare con la percezione stereoscopica delle salienze di un ambiente intenzionale.

* *Ricercatore del Centro per le Scienze Religiose della Fondazione Bruno Kessler di Trento. Mail*: pacosta@fbk.eu.
Il presente testo è stato letto per la prima volta al workshop «Neuroni artificiali e biologici: etica e diritto» (Trento, 3-4 dicembre 2020). Colgo qui l'occasione per ringraziare gli organizzatori dell'evento, Lorenzo Pavesi e Carlo Casonato, per l'invito a partecipare a un seminario genuinamente multidisciplinare.

Per «percezione stereoscopica delle salienze» intendo una visione non distaccata del contesto di azione che inclina il soggetto verso un certo giudizio o una certa scelta, come quando diciamo a qualcuno: «Ma non vedi come l'ha trattata: è intollerabile!»; oppure: «Non possiamo stare qui con le mani in mano: il livello di disuguaglianza nelle nostre società è nauseante!». Siccome non ritengo che il punto di vista morale possa retrocedere al qua di questo livello basilare di coinvolgimento, non interpreto queste descrizioni assiologicamente non neutrali (o *value-laden*, come si direbbe più concisamente in inglese) come dei *bias*, ma le considero come il succo stesso dell'esperienza morale.
Entro questo orizzonte teorico, per farla breve, agire moralmente non significa quindi, in primo luogo, isolare ragioni convincenti e ordinarle sulla base di sequenze inferenziali ineccepibili, ma avere una relazione ricca e risonante con il mondo circostante, che è allo stesso tempo un oggetto di attenzione e trasformazione, cura e rigetto, contemplazione e manipolazione, ammirazione e sgomento.

2. Paure sintomatiche

Fatta questa premessa, quello che intendo fare nello spazio a mia disposizione è proporre un breve esercizio di articolazione o riarticolazione di alcune intuizioni morali che circolano nella discussione intorno all'Intelligenza Artificiale. Mi accingo a questo compito con la determinazione di chi si dispone a remare contro corrente perché il modello epistemologico dominante praticamente in tutte le discipline scientifiche oggi fatica a capire che cosa vi sia di propriamente scientifico in una simile prestazione riflessiva.
Nella filosofia del Novecento esiste tuttavia una minoranza agguerrita di filosofi morali, forse sarebbe meglio dire di filosofe morali (da Hannah Arendt a Simone Weil, da Mary Midgley a Martha Nussbaum, da Elizabeth Anscombe a Cora Diamond), che ha sposato l'idea, riassunta alla perfezione da Iris Murdoch nella *Sovranità del bene*, secondo cui «la filosofia spesso consiste semplicemente nel trovare il contesto adatto per dire l'ovvio» (*philosophy is often a matter of finding a suitable context in which to say the obvious*)[1].
Il compito che mi prefiggo, insomma, è provare a formulare in maniera interessante qualcosa di scontato. Per farlo, prendo le mosse da un modo diffuso di vivere i progressi dell'*AI Technology*. L'atteggiamento in questione è caratterizzato da una dose non allarmante ma persistente di timore, apprensione, inquietudine rispetto alla possibilità che macchine costruite da esseri umani simulino alla perfezione o magari persino superino in efficienza l'intelligenza umana.
Descrivendo questo timore come «persistente ma non allarmante» vorrei richiamare l'attenzione sul suo carattere flessibile, malleabile, per certi versi adattivo. Voglio dire, il timore di cui parlo può magari presentarsi a un certo punto sotto forma di scetticismo circa la possibilità che un computer sia veramente in grado di battere una campionessa di scacchi o di poker, ma né svanisce né si acuisce nel momento in cui un nuovo computer effettivamente sconfigge la suddetta campionessa. Oppure l'inquietudine può riattivarsi di fronte alla prospettiva della creazione di un cervello cyborg, ma ciò non esclude che chi la prova possa anche destreggiarsi con disinvoltura tra le due immagini assiologicamente discordanti di un impianto congegnato per riabilitare delle funzioni umane essenziali e di un

[1] I. MURDOCH, *The Sovereignty of Good*, London 1970, 33.

esperimento alla Frankenstein che si nutre del sogno di una trasformazione fantasmagorica della natura umana.

3. Intelligenze rivoluzionarie

Che cos'è allora che alimenta questi timori e li rende così resistenti anche di fronte ai progressi tecnologici?
La mia impressione è che il fulcro di tali preoccupazioni, che personalmente condivido, non sia tanto il superamento od offuscamento del confine tra naturale e artificiale. I progressi della medicina moderna hanno ampiamente dimostrato fino a che punto le persone siano capaci di adattarsi a nuovi orizzonti di possibilità pratiche senza stravolgere i propri quadri di riferimento morali.
La questione è un'altra e la formulerei così. Per molte persone il fatto di essere creature intelligenti ha poco a che fare con specifiche prestazioni cognitive, ma con un'intuizione generale tanto vaga quanto robusta dal punto di vista doxastico. L'idea a cui faccio riferimento, riassunta in poche parole, è che nelle vite umane sia in gioco qualcosa di speciale e per «speciale» intendo dotato di un'importanza fuori dall'ordinario.
Per chiarire questo punto sono disponibili vocabolari molto differenti. Io mi servirò qui di quello tradizionale dei beni «architettonici».
Prendiamo il caso della genitorialità. Ridotto all'osso, essere padri consiste effettivamente nello svolgere più o meno bene una serie di mansioni educative e di cura. Ma il significato dell'essere «padri» non è di certo traducibile in una tabella che sommi l'efficienza, l'utilità o la dispendiosità di tali compiti. C'è qualcosa che allo stesso tempo si svela e si vela nell'essere genitore e questo «qualcosa» funziona come un'asse attorno a cui finiscono per ruotare più o meno ordinatamente le esperienze di una vita intera.
L'intelligenza è esattamente la capacità di avere accesso a tale esperienza rivelatoria in una forma che è simultaneamente critica e ricettiva e che si manifesta secondo forme espressive variegate che possono andare dall'epifania poetica all'umorismo (del quale conta qui, non tanto la «logica», quanto la capacità di riconfigurare la percezione della realtà mediante l'apprensione istantanea delle sue incongruità).
Ma il punto non è solo l'effetto di *disclosure*, di rivelazione. Il succo della «specialità» a cui ho fatto cenno sopra è piuttosto la sensazione robusta e persistente che l'intuizione che la veicola non sia tanto il motore di un processo di apprendimento selettivo e distaccato quanto piuttosto la condizione di possibilità di una *trasformazione* profonda dell'esistenza. Questa può assumere persino la forma di una conversione a U grazie alla quale viene rivoluzionata la morfologia stessa del proprio essere al mondo, dove per «morfologia» intendo cose come il «senso», l'«atmosfera», la «cornice», la «centratura» dell'agire e del pensare.

ISSN 2284-4503

4. Un nuovo imperativo categorico

Le persone che monitorano i discorsi sull'Intelligenza Artificiale con un'attenzione che può talvolta sfociare in apprensione lo fanno in genere perché hanno a cuore questa esperienza limite e la interpretano come un tratto distintivo dell'intelligenza umana. Magari poi la rinominano «saggezza» o «sapienza» in omaggio a una lunga tradizione, non solo occidentale, ma quello che hanno in mente è comunque una forma di perspicacia che, più che come un algoritmo, opera come un punto di sintesi creativa e contingente dei molti modi in cui si manifesta l'intelligenza nel mondo, che vanno ben al di là di ciò che avviene tra i congegni delle macchine computazionali o all'interno di un cervello umano (i 1200 grammi di paté elettrificato di cui parla David Foster Wallace in *Tutto e di più* o il chilo e mezzo di spugna imbevuta di sangue, «la colazione di un cane», nominata da Kurt Vonnegut in *Cronosisma*)[2].

Potrei sbagliarmi, ma ho l'impressione che dal riconoscimento di questa esperienza che ho provato qui a descrivere con le mie parole e le mie categorie filosofiche dipenda il riconoscimento del nucleo normativo non solo della forma di vita moderna, ma di tutte le grandi civiltà e tradizioni sapienziali.

Per riassumere il mio ragionamento in pillole direi allora che la paura persistente ma non allarmata di fronte ai progressi dell'Intelligenza Artificiale dipende soprattutto dal timore per i suoi possibili danni collaterali in questo ambito centrale dell'esperienza umana: dal turbamento, cioè, di fronte alla prospettiva di una perdita significativa nel *potenziale trasformativo* dell'intelligenza umana.

Il problema, se vogliamo, non è tanto l'insulto alla natura né la perdita di un'autonomia che, se esiste, può manifestarsi solo in condizioni ideali, quanto lo spettro di una meccanicità, a suo modo umoristica, in cui il carattere caleidoscopico e risonante dell'intelligenza umana venga ridotto alla linearità del calcolo. D'altra parte, non è certo alle capacità di computazione che ci riferiamo quando ci chiediamo più o meno ritualmente se usciremo migliori, ossia più intelligenti, dall'emergenza sanitaria che stiamo vivendo da alcuni mesi.

In conclusione, per venire incontro a chi, a dispetto del mio *caveat* iniziale, sente comunque il bisogno di una regola a cui affidarsi per fugare i timori circa le conseguenze della rivoluzione dell'Intelligenza Artificiale, suggerirei di adottare questa riformulazione dell'imperativo categorico kantiano: «Nel tuo lavoro agisci in modo che la massima che orienta il tuo pensiero non pregiudichi aprioristicamente la possibilità di riconoscere quella visione trasformativa dell'intelligenza che per molte persone ragionevoli è una condizione di intelligibilità della loro esistenza».

[2] Cfr. D.F. WALLACE, *Tutto, e di più. Storia compatta dell'∞*, trad. it., Torino, 2005, 20-21; K. VONNEGUT, *Cronosisma*, trad. it., Milano, 1998, cap. 27.

ISSN 2284-4503