

Questo documento riporta il testo della postfazione poi pubblicata in Franco SARCINELLI, *Essere umano. Per un'etica del ben-essere*. Mimesis, Milano 2024, pp. 317-323.

Le domande dell'essere umano e le risposte dell'Intelligenza Artificiale

Fabio Fossa,
Dipartimento di Ingegneria Meccanica
Politecnico di Milano
fabio.fossa@polimi.it

In un bel passaggio de *Il Principio Responsabilità*, recentemente ripreso dalla filosofa della tecnologia Shannon Vallor,¹ Hans Jonas lamentava l'impreparazione del suo tempo alle questioni che era chiamato ad affrontare, scrivendo: "abbiamo tanto più bisogno della saggezza quanto meno crediamo in essa".² Il bisogno di riconquistare uno spazio autentico per la saggezza, e di ritrovare fiducia nell'atteggiamento teorico e pratico che lo abita, vibra nelle pagine di Jonas e Vallor così come in quelle di Sarcinelli. Allo stesso modo, è condivisa la realizzazione che, se un senso ancora rimane aperto per quanto la tradizione filosofica occidentale indica con l'onorevole nome di saggezza, esso sia ora (di nuovo?) inestricabile dal sapere delle scienze tecnologiche.

L'apertura del pensiero sul ben-essere umano – e, con esso, dell'universo che lo ospita – ai saperi propri del conoscere scientifico rivela la preoccupazione epistemica di chi si pone le domande dell'etica di fronte a un mondo la cui complessità sgomenta e ammutolisce. Al cospetto delle immagini frastagliate del presente, e delle visioni opache del futuro, il bisogno disperato di vederci un po' più chiaro è forse tra le cifre più significative delle sempre nuove forme in cui la finitezza umana si rivela nelle pieghe della storia.

La presa di consapevolezza dell'inseparabilità di saggezza pratica e conoscenza scientifica assume anch'essa forme nuove e diverse. Se da una parte si fa strada, pur nelle innumerevoli e innegabili difficoltà che la accompagnano, l'esplorazione di esperienze interdisciplinari volte proprio a stimolare l'integrazione dei saperi, dall'altra è ben visibile la tendenza a cercare le risposte che ci mancano là dove continuano ad arrivare, a ritmi sbalorditivi, le migliori risposte a domande che a volte nemmeno si credeva di avere. Se è un difetto di conoscenza a marcare le angosce di un'umanità sempre più esposta alle conseguenze della propria imprevidenza, non può che sorgere la tentazione di affidare il compito di colmare la lacuna – e, con questa, almeno in parte, anche la nostra stessa

¹ Vallor, S. *Technology and the Virtues. A Philosophical Guide to a Future Worth Wanting*. Oxford University Press, 2016, p. 29.

² Jonas, H. *Il principio responsabilità. Un'etica per la civiltà tecnologica*. Einaudi, 2009, p. 29.

finitezza – proprio al frutto più straordinario del nostro saper fare. Perché non *produrre* le risposte che ci servono, che non sappiamo più darci, che non ci fidiamo più a formulare insieme?

Non sorprende che dai sistemi di Intelligenza Artificiale si inizi ad attendersi lumi anche su quell'insieme di questioni che fanno capo, in un modo o nell'altro, alla radicale domanda sulla vita buona. Se si è stati capaci di costruire sistemi in grado di elaborare risposte sorprendentemente adeguate a domande ben determinate e circoscritte, è solo – si può pensare – una questione di tempo: avremo presto a disposizione sistemi che rispondano a domande meno circoscritte, ma assai più importanti. È il momento di realizzare sistemi che sappiamo soddisfare l'esigenza così profondamente avvertita di vederci più chiaro, di tracciare una rotta sicura che tragga la nave del mondo fuori dalle acque agitate in cui ora scricchiola e beccheggia.

Che il successo odierno dei sistemi di Intelligenza Artificiale risieda nella loro capacità di svolgere compiti complessi senza richiedere costante intervento e supervisione umana è d'altronde difficile da negare. Non tutti i compiti, però, possono essere delegati allo stesso modo. Uno sguardo anche d'insieme alla pervasività e trasversalità delle tecnologie di IA coglie immediatamente l'evidenza degli effetti multiformi che seguono alla loro progettazione e al loro uso. Se, da una parte, i sistemi di IA offrono supporti validi per la raccolta e l'elaborazione di informazioni altrimenti inaccessibili, la questione della vita buona, o dei fini e dei valori da perseguire (anche tramite questi sistemi stessi), non ne viene in alcun modo "ottimizzata". Al contrario, si può ben dire, la questione si complica, rendendo ancora più urgente districare l'intreccio di prassi scientifiche, mediazione tecnologica, e valori morali che costituisce il tema centrale dell'etica della tecnologia.

Mettere a fuoco la questione dei valori posta dalla diffusione di sistemi di IA è utile per sollevare il velo illusorio della delegazione dalla domanda sul ben-essere di cui si è detto sopra, e a cui il libro di Sarcinelli approda. Nel caso di questa domanda – una domanda che non è come le altre – la tecnologia non può far altro che restituirci quello che siamo. O meglio, non può far altro che restituirci le risposte più o meno consapevoli di chi le tecnologie di IA le concepisce, progetta, sviluppa, commercia, gestisce, usa. E se le risposte alla domanda sulla vita buona che l'IA può dare sono inevitabilmente umane, dell'umano hanno tutte le qualità ben note. Diversi attori sociali nutrono diverse convinzioni circa quali siano i valori, le finalità e le intenzioni che sia bene o lecito perseguire con e grazie agli strumenti tecnologici. Ma valori, intenzioni e finalità possono differire anche in modo sostanziale tra i diversi gruppi sociali e tra i diversi individui coinvolti. E non è certo un output di un sistema di IA a poter ricomporre differenze di tal genere – essendo *queste* differenze a determinare la bontà dell'output e, di conseguenza, l'efficienza e l'affidabilità del sistema che lo emette.

L'essere umano rimane inevitabilmente al centro del dramma, e la sua tecnologia – per quanto in modo assai più mediato – rimane un semovente animato da vita riflessa. Non è lecito aspettarsi nessuna decentralizzazione dell'umano dalla delegazione delle domande sul ben-essere a sistemi di IA. Quanto ci si deve aspettare, al contrario, è un *surplus* non delegabile di centralità umana, che si trova ora *anche* a dover fare i conti con l'inedito impatto dei sistemi di IA sul delicato e (a causa nostra) precario ecosistema di cui siamo parte.

Si intuisce il motivo per cui la questione ecceda il punto di vista meramente tecnico. Posto che l'obiettivo della ricerca tecnologica sia di fornire strumenti efficienti, la questione del ben-essere non può essere soddisfatta studiando i modi in cui un determinato compito (vivere bene?) può essere eseguito nel minor tempo, con il minor dispendio di risorse e nel modo più accurato possibile. L'intrico sociale in cui l'IA è inserita rende necessario volgere l'attenzione anche all'impatto del suo uso e del suo funzionamento sui valori pertinenti al suo ambito di applicazione. È necessario prendere in considerazione gli effetti che una tecnologia esercita su ciò che sta a cuore agli esseri umani interessati, su ciò a cui essi danno valore, su ciò che ritengono degno di approvazione o di biasimo. L'efficienza dell'IA non è solo una questione di precisione e risparmio, ma mette in gioco un insieme di fattori assai più ampio che ha a che vedere con ciò che gli esseri umani coinvolti, in modo attivo o passivo, nella sua realizzazione e nel suo uso ritengono sia bene e male. Aspettarsi da questa indicazioni utili su dove dirigere lo sguardo per riflettere in modo più profondo e radicale sui caratteri e le prassi del ben-essere è certamente lecito; aspettarsi che rifletta al posto nostro sul nostro stesso ben-essere è insensato, essendo quest'ultimo il banco di prova dell'efficienza della prima.

Diventa un po' più chiaro perché abbiamo bisogno di un'indagine *etica* sul ben-essere che *anche* includa l'IA – cioè di una discussione critica che fissi lo sguardo non solo sull'efficienza delle nostre tecnologie, ma anche sul bene del genere umano nella complessa congiuntura storica che ci troviamo a vivere. Il bisogno di saggezza rimane, così come però rimane l'esigenza di vederci più chiaro su come esercitarla nella prassi – esigenza che difficilmente può essere soddisfatta evitando il dialogo con i saperi delle scienze tecnologiche e le prestazioni dei sistemi di IA. Solo il concorso e la contaminazione reciproca dei saperi umani possono aiutarci a tracciare quella rotta di cui in modo sempre più netto avvertiamo il bisogno.

Invece che rendere obsoleta la domanda sul benessere dell'uomo e dell'universo di cui è parte, quindi, l'IA ne rende ancora più urgente la posizione e l'esplorazione radicale. Un quadro valoriale il più possibile coerente e chiaro è necessario se vogliamo che l'attività tecnologica di ricerca e sviluppo sia eticamente consapevole e venga praticata alla luce di tale consapevolezza. Se, al contrario, la riflessione sul quadro etico viene trascurata, si rischia di condannare all'inefficacia lo sforzo morale, viziato dall'inevitabile emergere di contrasti tra valori ugualmente rilevanti, zone grigie, discorsi

generici difficili da collegare ai casi reali e strumentalizzazioni da parte di ben più potenti forze sociali.

L'impreparazione di cui parla Jonas rimane ben visibile. L'idea di lavorare insieme a un quadro valoriale che orienti la progettazione e l'uso dell'IA è apparsa a molti come una chimera o un residuo di un modo di pensare inattuale, da guardare con sospetto. Molti dubitano del fatto che lo sforzo di sistematizzare i valori a cui adeguare l'uso dell'IA sia compatibile con la pluralità delle ragionevoli opinioni umane sul tema. E tuttavia, per quanto simili sospetti siano senz'altro parzialmente condivisibili, il bisogno di porre la domanda del ben-essere umano e della relativa saggezza non sembra soddisfabile altrimenti che con un pensiero su ciò che reputiamo buono e ci sta a cuore in un mondo caratterizzato da livelli altissimi di complessità.

E forse è bene qui anche regolare le aspettative che si nutrono nei confronti dell'impresa. Tracciare un quadro valoriale non significa gettare le basi per una procedura in grado di risolvere ogni contrasto in modo oggettivo e inappellabile. Tantomeno è onesto trasformare un quadro valoriale in una tavola dei comandamenti. I quadri sono mappe, carte che possono servire ad orientare i comportamenti – a tracciare rotte. I quadri generali, se ben intesi e ben usati, non riducono qualsiasi riflessione a ciò che sembra buono a un gruppo di persone in un angolo di mondo. Lo scopo è esattamente l'opposto: offrire a chi sente l'esigenza di confrontarsi con il problema una serie di linee guida per sollevare, discutere e affrontare il problema del ben-essere – non per scavalcare l'atto del giudizio, quindi, ma per innescarlo o stimolarlo e, così, fare filosofia del ben-essere.

Se è un ritorno alla frequentazione del pensiero e della prassi della saggezza che occorre coltivare, la responsabilità del compito non può che poggiare sulle spalle dell'essere umano. L'antropocentrismo è giustamente individuato da più parti come il paraocchi che impedisce l'affermazione di una visione del mondo più adeguata alle sfide della contemporaneità e alla posizione dell'essere umano nel cosmo. Il portato teorico della sua revisione – di cui appena si stanno compiendo i primissimi passi – è assolutamente degno di nota. Ma quale ne è il correlato pratico? Cosa significa decentralizzare l'umano in vista del compito d'azione che si profila?

Se la posizione di centralità normativa dell'umano è da decostruire, la posizione di effettiva centralità pratica dell'agire umano è da assumere in tutta la sua vertiginosa pregnanza. L'onesta e disillusa (persino impietosa) diagnosi del peso dell'agire umano nel mondo è fondamentale per la piena realizzazione della portata del compito da svolgere, in tutte le sue innumerevoli sfaccettature. Bisogna farsi carico di un antropocentrismo della responsabilità che riconosca senza sconti – e senza incriminazioni di comodo del naturale o dell'artificiale – quanto del ben-essere nostro e altrui dipenda dai modi e dalle determinazioni della nostra esistenza, materiale e spirituale. Se il gesto futuro

dell'umano deve essere un gesto di cura, non ci si può esimere dal riconoscere il ruolo effettivo che giochiamo nel nostro ecosistema e le responsabilità che ne conseguono.

Sconti di responsabilità, si è detto, non possono venire né dall'addossare al mondo le conseguenze delle nostre azioni, né dall'indicare i sistemi di IA come oracoli di salvezza o agenti di sventura. Dietro la complessità della vicenda dell'IA sta un elemento costante, che bisogna abituarsi a non perdere di vista: l'essere umano e l'eterno suo compito, assunto o disatteso, di vivere bene con sé, con gli altri viventi, con la natura di cui è parte. La centralità dell'umano non riguarda solo i valori a cui l'etica dell'IA si sforza di allineare le tecnologie – i quali non possono che essere i nostri valori, della cui traduzione pratica siamo ineluttabilmente responsabili (che lo si riconosca o meno). Riguarda allo stesso modo le intenzioni, i bisogni, le necessità, i rapporti di potere che, tramutati in prodotto tecnologico, diventano materiali e strutturano il mondo. A ciò si aggiunga la logica interna di costante accrescimento orizzontale e verticale che l'avanzamento tecnologico esibisce, camuffando la sua natura strumentale per proporsi, con il concorso assiduo delle forze sociali che ne profittano, come fine in sé stesso. Se ciò accade, i valori e le esigenze che le tecnologie stesse sono state create per realizzare scivolano paradossalmente in secondo piano, con il rischio di confondere il ben-essere umano con l'auto-asservimento ai propri stessi prodotti – ovvero, a chi da essi ne profitta.

Fare dell'IA una forza benefica, uno strumento per la gestione saggia dei problemi contemporanei e non un'arma di discriminazione, sperequazione, manipolazione e controllo è un compito che non può essere a sua volta delegato e il cui risultato dipende interamente dall'impegno che in esso si ripone. Questa la sfida notevole dell'etica dell'IA. Una sfida che conduce sulla soglia di domande smisurate – cosa vogliamo, cosa ci sta a cuore, cosa è bene e male, in che mondo vogliamo vivere, cosa dobbiamo agli altri viventi – che forse non siamo più così abituati a porre, ma sulle quali dobbiamo al più presto tornare a riflettere; e alle quali, in certa misura, dobbiamo reimparare a rispondere. L'articolato discorso che Sarcinelli declina nelle pagine del suo libro, di respiro ben più ampio rispetto al problema affrontato nelle mie poche pagine, accetta la sfida a viso aperto e con onestà filosofica, invitando i lettori a un modo di fare filosofia di cui si avverte chiaro il bisogno.