

L'evoluzione degli esseri umani e delle tecnologie: una proposta per il biodiritto

Amedeo Santosuosso

Professore straordinario di filosofia del diritto presso Scuola Universitaria Superiore IUSS Pavia. Mail: amedeo.santosuosso@iusspavia.it.

La questione del rapporto tra esseri umani e macchine è sempre più al centro del dibattito filosofico e, probabilmente, dovrebbe essere una delle missioni principali del biodiritto e di *BioLaw Journal*. Su questo punto è centrato il mio intervento.

Uno spunto interessante è fornito dai robot cognitivi, che possono essere utilizzati in ambito assistenziale o in fabbrica. Queste macchine processano un'enorme quantità di dati su canali diversi, prelevano dati dall'ambiente mediante sensori e ne estraggono informazioni; comunicano con altri sistemi informatici per quanto riguarda il riconoscimento delle immagini (per es. l'identificazione corretta di un paziente a mezzo di riconoscimento facciale), dialogano con gli umani attraverso il linguaggio naturale (usando l'IA generativa). L'elaborazione dei dati, anziché avvenire in locale, all'interno del robot, può far capo a server diversi, dislocati in infrastrutture diverse, in ragione del tipo di dati da processare. Nessun umano oggi può essere in grado di elaborare in tempo reale questa enorme quantità di dati, e neppure di controllarne, nel dettaglio, l'elaborazione¹.

Di fronte a questo tipo di problemi vengono spesso usate alcune formule, che paiono più che

altro delle forme di esorcismo. Senza disconoscere in alcun modo la serietà delle questioni in gioco, provo piuttosto a uscire da alcuni percorsi argomentativi che paiono ripetitivi e ad affrontare il centro dei problemi più importanti.

Per esempio, ha sicuramente valore l'ispirazione di idee come tenere l'umano nell'ambito della decisione (*the human in the loop*, letteralmente «l'umano nel circuito»), oppure di umanesimo digitale oppure di intelligenza artificiale antropocentrica (*human centered artificial intelligence*) o l'idea che la tecnologia funzioni da "copilota" senza sostituire l'umano, ma non si può non rilevare che spesso queste idee, al netto di alcuni espedienti commerciali, appaiono più come un desiderio, una rassicurazione, un'opzione culturale/emotiva, che magari è condivisibile (chi vuole che l'umano sia escluso dal processo decisionale?), ma che, in mancanza di un adeguato esame di quello che realmente accade e, soprattutto, di un'indicazione dei mezzi e dei passaggi tecnici e organizzativi per raggiungere quel risultato, può essere fonte di frustrazione ulteriore.

Tipico il caso della formula *the human in the loop*, che è poco efficace se non ci si fa carico di dire quali siano i confini del *loop*, cosa accada al suo interno, quale sia la posizione che in quest'area si vuole riservare all'umano. Essere nel giro, d'accordo, ma come? In quale posizione? Con quale controllo dei dati e delle decisioni? Con quali poteri, specie in caso di divergenza dai suggerimenti delle macchine?

E allora, bisogna chiedersi cosa significhi la centratura sull'umano o che l'umano sia nel *loop*. Si deve pensare che la centratura sull'umano e l'essere l'umano *in the loop*, se non sono un mero

estesi. Sul tema dei robot cognitivi è in corso un'interessante collaborazione scientifica tra la società *Over-sonic*, creatrice del robot Robee, e il Centro di ricerca *International Center on Law Science and New Technologies* (ICLT) presso la Scuola IUSS di Pavia.

¹ Le riflessioni contenute in questo mio intervento traggono spunto dal lavoro fatto con Giovanni Sartor in preparazione del volume A. SANTOSUOSSO, G. SARTOR, *Decidere con l'IA*, Bologna, 2024, al quale rinvio per un'argomentazione e riferimenti bibliografici più

richiamo simbolico, facciano riferimento a un'effettiva capacità di un controllo sulle finalità perseguite (autonomamente o collaborando con gli umani) dal sistema, e sugli effetti collaterali che ne risultano. Questa è un'affermazione più realistica che merita un approfondimento.

La centratura sulle finalità, in particolare, ci interroga su chi e come definisca le finalità e, poi, su come esse si realizzino e su come avvenga il controllo circa il fatto che quelle finalità siano effettivamente rispettate e siano conformi alle nostre intenzioni. Nel caso dei robot cognitivi, per esempio, se non si vuol cadere nella contraddizione di rimpiazzare l'umano, che non riesce a processare quella grande quantità di dati di cui si è detto prima, con un sistema artificiale che svolga quella funzione, ampliando così il problema, è necessario immaginare qualcosa di diverso, che garantisca il rispetto di quella finalità di cui si è detto.

Ma allora, se è necessario trovare una forma di controllo e interazione tra l'umano che produce la tecnologia, le finalità intrinseche di tecnologie che superano l'umano nella capacità di elaborazione delle informazioni e il controllo dell'effettivo perseguimento di quelle finalità, alla fine è necessario rendersi conto che si sta realizzando un cambiamento della natura dell'umano che interagisce con questi strumenti. A questo punto la domanda ulteriore diventa: di quale umano si sta parlando? È un umano fisso nel tempo e nello spazio? O come altrimenti definito?

Dopo Darwin non dovrebbe esservi dubbio circa la non fissità creazionista di cosa sia umano. A ben vedere, anche in assenza di evoluzione biologica e accanto ad essa, nuove condizioni sociali e tecnologiche possono cambiare profonda-

mente l'essere umano, e la nostra concezione di esso. I dialoghi tra Fitzroy, comandante del brigantino Beagle, e Darwin, che, entrambi dalla prospettiva di credenti cristiani, si interrogano su se abbiano o meno la qualità di umani gli abitanti della Terra del fuoco (nudi a temperature polari, ricoperti di lunghi peli, con pratiche sessuali ai loro occhi animalesche ecc.), così poco simili all'umano europeo sul quale entrambi avevano costruito la loro idea di umano, sono illuminanti e di un interesse straordinario².

Bisogna allora assicurarsi che l'essere umano dell'era dell'IA sia veramente un umano con accresciute capacità di conoscere e di agire, frutto dell'integrazione delle proprie competenze cognitive e pratiche con quelle della macchina, piuttosto che un umano indebolito e dequalificato (*deskilled*), dopo essere stato sostituito dalla macchina nella sua autonomia e nella sua dimensione lavorativa e sociale. La logica sembra essere la seguente: per avere una IA centrata sull'umano è necessario che l'IA abbia anche l'obiettivo di abilitare l'umano e che l'umano potenzi alcune sue capacità, perché, ove rimanesse fermo, non avrebbe possibilità di essere al centro di questo straordinario sviluppo fornito dall'IA. Se questo è il caso, la domanda diventa *chi* sarà il soggetto, e quali caratteristiche avrà il *chi*, di questa trasformazione.

In breve, la via sembra essere quella di una coevoluzione di umani e macchine, nel contesto di una collaborazione che non è di pari natura e grado e che, in ogni caso, richiederà agli esseri umani di adattarsi a un mondo nel quale la nostra ragione non rappresenta il solo modo di conoscere la realtà o di muoversi al suo interno³. Il

eliminazione dei fuegini nulla toglie alla sincerità e profondità del dibattito tra Fitzroy e Darwin.

³ A. SANTOSUOSSO, *The human rights of nonhuman artificial entities: An oxymoron?*, in *Yearbook of Science*

² La ricostruzione è in H. THOMPSON, *Questa creatura delle tenebre*, Roma, 2021 (prima edizione inglese 2005). Che qualcuno possa avere visto nelle teorie di Darwin una legittimazione delle successive pratiche di

futuro ci riserva probabilmente l'avvento di altre tecnologie e a un passo probabilmente ancora più rapido. A fronte di ciò una risposta fondamentale sarà quella della conoscenza dei fenomeni e della formazione. Tuttavia, ed è un punto cruciale, i problemi che ci troviamo a fronteggiare non possono essere risolti in termini di puro potenziamento individuale, che, da solo, sarebbe foriero di ulteriori discriminazioni. È necessaria una discussione pubblica su come ed entro quali limiti e secondo quali modalità si voglia usare l'IA nelle decisioni.

E anche il quadro normativo va in alcuni punti ripensato nella sua stessa ispirazione. Nota giustamente Sergio Quintarelli che l'AI Act è poco incisivo in alcuni punti, per esempio laddove non prevede forme di *redress by design*, cioè «l'idea di stabilire, fin dalla fase di progettazione, dei meccanismi che garantiscano la ridondanza, sistemi alternativi, procedure alternative di valutazione per poter efficacemente rilevare,

controllare, correggere le decisioni sbagliate prese da un sistema perfettamente funzionante»⁴. Oppure un meccanismo contro il pregiudizio a favore dell'automazione (*automation bias*), che è la tendenza a conformarsi con la previsione della macchina.

È necessario un piano di investimenti sul fattore umano, di educazione, dalle scuole (dove è necessario rompere quel circolo vizioso per cui i giovani hanno bisogno di essere formati da insegnanti che a loro volta, però, non sono in grado di formarli, per problemi di *forma mentis* e di mancanza di conoscenze), alle università, alla formazione professionale dei giudici e degli amministratori e dei cittadini in generale

Ecco una missione che i cultori del biodiritto e *BioLaw Journal*, a dieci anni dalla sua fondazione, potrebbero, e oserei dire dovrebbero, assumere su di sé per il prossimo futuro.

and Ethics/Jahrbuch für Wissenschaft und Ethik, 19, 1, 2015; ID., *About coevolution of humans and intelligent machines: Preliminary notes*, in *BioLaw Journal – Rivista di BioDiritto*, S1, 2021, Si veda anche E. ASHFORD

LEE, *The coevolution: the entwined futures of humans and machines*, Cambridge, 2020.

⁴ S. QUINTARELLI, *Intelligenza artificiale*, Torino, 2020, 131.