

INTRODUZIONE. INTELLIGENZA ARTIFICIALE, ETICA E SOCIETÀ

LUCIA GALVAGNI

In questo particolare momento storico le tecnologie digitali e l'Intelligenza Artificiale rappresentano sfide e temi centrali nel dibattito non solo tecnico e scientifico, ma anche sociale e politico, suscitando entusiasmi e stupori e generando al medesimo tempo paure e timori rispetto al loro possibile impatto e alle evoluzioni presunte e inattese che potrebbero avere. Per questo pare importante raccogliere spunti di riflessione da quegli ambiti di studio che tradizionalmente si sono interrogati sui fenomeni e sulla loro portata, interpretandoli alla luce del significato che essi possono avere e insieme per l'impatto che la loro introduzione può implicare rispetto alla ricerca e alla costruzione di senso che caratterizza l'esistenza e la vita degli esseri umani. Ci si può chiedere, ancora, quali siano e quali potrebbero essere gli effetti dell'uso di queste tecnologie per le persone e per le dinamiche interpersonali, sociali e politiche, così come nei termini delle evoluzioni possibili delle nostre società.

Dal momento che da essa dipende la progettazione, la realizzazione e la messa in opera di tali tecnologie, la dimensione tecnico-scientifica sembra rappresentare la voce prevalente nello scenario attuale e per questo essa sembrerebbe giocare il ruolo più importante: quale ruolo compete allora e ancora alle scienze umane e che contributo possono portare queste discipline nella riflessione sulle finalità da dare alle tecnologie, sui modelli ai quali ispirarle e sulle modalità per utilizzarle? Come la filosofia, la storia, il diritto, le scienze umane e sociali entrano nel dibattito e quali strumenti offrono non solo per capire e interpretare, ma

anche per prevedere e anticipare scenari che possono risultare più desiderabili e auspicabili, più equi ed egualitari, più rispettosi e sostenibili?

Per entrare in questo esercizio riflessivo ci è sembrato importante recuperare alcune nozioni etiche, politiche e più ampiamente filosofiche che possono orientare nella lettura e nella comprensione degli scenari e dei contesti che il digitale e l'artificiale sembrano profilare e definire oggi. In questa sezione degli "Annali di Studi Religiosi", intitolata *Intelligenza Artificiale, etica e società*, abbiamo raccolto per questo i contributi di alcuni studiosi che si sono interrogati in merito: da essi emerge una riflessione inter- e multi-disciplinare sulla realtà dell'Intelligenza Artificiale e sul ruolo che le tecnologie digitali giocano alla luce del loro valore epistemologico, del loro impatto etico, sociale e politico, del loro possibile contributo rispetto all'esistenza umana.

I temi sui quali si riflette e si dibatte così ampiamente non sono nuovi, basti pensare che in un suo famoso articolo del 1950 Alan Turing osservava che se la domanda "Le macchine sono in grado di pensare?" poteva sembrare insensata in quel particolare momento storico, in prospettiva, ad esempio alla fine del secolo, non sarebbe probabilmente suonata così impertinente (Turing 2024). A 75 anni di distanza da quelle riflessioni lo scenario sul quale ci confrontiamo e dibattiamo sembra essere proprio questo: le macchine non solo simulano — come si era immaginato e preventivato in passato in importanti convegni e conferenze —, ma esse a propria volta producono, pensano, interagiscono, mettendo in atto forme di esercizio della razionalità e della creatività attribuite e considerate come specificamente, sinora forse esclusivamente, umane (Cortesi 2025).

I saggi raccolti nella presente sezione portano lo sguardo su questioni diverse che l'Intelligenza Artificiale (IA) solleva e su alcune sfide che ci pone di fronte. I primi due contributi rappresentano un dialogo tra filosofia e informatica: James Brusseau riformula e rilancia quesiti antichi per meglio inquadrare e capire gli scenari dell'IA, mettendo in luce come essa permetta di rileggere questioni importanti, tradizionalmente dibattute dalla filosofia; un confronto a più voci, frutto di una tavola rotonda che ha coinvolto filosofi — Paolo Costa, Eugenia Lancellotta e Boris Rähme — ed informatici — Michela Milano e Paolo Traverso — della Fondazione Bruno Kessler mette a tema il valore delle credenze

e delle convinzioni presenti nelle nostre società rispetto alla realtà e alle applicazioni dell'Intelligenza Artificiale. Gli altri tre saggi della sezione pongono al centro e discutono importanti questioni etiche correlate all'uso dell'IA: Marta Fasan, in un contributo di taglio giuridico, considera il tema dell'equità e dell'eguaglianza alla luce delle applicazioni dell'Intelligenza Artificiale; Eugenia Lancellotta riflette sull'uso e l'applicazione delle tecnologie digitali in medicina, analizzando l'ambito e la cura della malattia mentale; Daniel Toscano Lopez, infine, entra nel merito della riflessione filosofico-politica e considera quale impatto etico e sociale le nuove tecnologie possono e potrebbero avere anche in una prospettiva globale.

Più in particolare, l'articolo di James Brusseau offre un approfondimento di tipo filosofico, nel quale quesiti antichi vengono riformulati per inquadrare e capire gli scenari che l'Intelligenza Artificiale crea, recuperando e problematizzando concetti quali l'identità e l'autenticità, la curiosità, la libertà e la creatività. Nella sua duplice competenza informatica e filosofica, Brusseau affronta nel merito ciascuna di queste nozioni e sottolinea quanto e come esse vengano interpellate dalle nuove tecnologie: come si definisce la convenienza, come si può ridefinire l'autenticità, interpretata secondo l'accezione heideggeriana e considerata rispetto agli scenari digitali, presenti e futuri? Come si potrebbe rimodulare la nozione di libertà e come le raccomandazioni fornite dagli algoritmi potrebbero tutelarne l'esercizio piuttosto che limitarla? L'Intelligenza Artificiale potrebbe modificare, in prospettiva, le dimensioni della curiosità e della serendipità? Per contrastare il rischio di omologazione e la possibile mancanza di variazioni si sono già identificati alcuni strumenti *ad hoc*, quali il *curiosity engine* e l'introduzione di raccomandazioni atipiche, che ricorrendo all'uso di filtri specifici permettono al sistema non solo di non precludere, ma al contrario addirittura di promuovere la libertà delle persone che utilizzano queste tecnologie.

Nella ricerca sull'IA, sottolinea Brusseau, si osservano al momento due linee principali, una prima che enfatizza in maniera quasi euforica il valore della convenienza e dell'autenticità digitale, una seconda orientata invece alla promozione della curiosità e della libertà intese in senso più creativo. Questi due diversi approcci riflettono anche le diverse

modalità di interpretare la questione dell'identità personale: lì dove l'identità personale venga intesa come possibilità di incontrare il proprio vero sé, la personalità potrebbe risultare valorizzata dal ricorso all'IA, favorendo la considerazione di interessi e inclinazioni della persona. D'altra parte, però, se si intende l'identità come qualcosa che si modifica e si definisce progressivamente, si potrebbe pensare a programmare gli algoritmi di raccomandazione ai fini di implementare cambiamenti in maniera provocatoria e imprevedibile, sollecitando ad esempio la curiosità verso spazi nuovi e realtà diverse. Le questioni etiche e filosofiche che si sollevano nel dibattito riguardano così l'identità sia da un punto di vista morale, che in termini giuridici: se ci può essere il desiderio di diventare qualcun altro, non solo nel senso di voler migliorare, ma anche e soprattutto nel senso di voler cambiare radicalmente il proprio modo di presentarsi e di proporsi, il ricorso all'Intelligenza Artificiale e agli algoritmi di raccomandazione potrebbe essere uno strumento per promuovere tale processo e attivare queste dinamiche.

Come si guarda oggi all'intelligenza artificiale, quali sono le convinzioni individuali e collettive e quali credenze sono presenti nelle nostre società rispetto a tali tecnologie? Come le nostre società vivono la crescente evoluzione digitale, che implica un'importante e impattante trasformazione?

In una riflessione sulle credenze e le convinzioni inerenti all'IA, i filosofi Paolo Costa, Eugenia Lancellotta e Boris Rähme hanno sollevato alcuni importanti quesiti, affrontati e approfonditi in un primo momento nelle scuole con gli studenti, e si sono poi confrontati in merito con Michela Milano e Paolo Traverso, informatici esperti di IA. Le "credenze resilienti" in merito all'IA, quelle che più sembrano venir messe in discussione, sono la nozione di intelligenza, il confronto tra intelligenza artificiale e intelligenza naturale/umana, il rapporto tra intelligenza artificiale e politica. Perché le convinzioni contano e sono importanti anche rispetto a come si progettano le tecnologie? Se si prende in considerazione il "senso comune", nel quale sono presenti anche le intuizioni e i bisogni delle persone, è possibile identificare importanti indicatori di quanto conta per loro e questo può già rappresentare un significativo insieme di credenze che fungono da riferimento nelle società, una sorta di "saggezza comunitaria". Qual è, quale può essere il

vantaggio di approcciare le nuove tecnologie tramite forme di saggezza comune, facendo spazio alla cosiddetta *common morality*? Creare un ambiente e un ecosistema dove l'Intelligenza Artificiale venga percepita come affidabile, *trustworthy* in inglese, può aiutare a muovere in direzione di concezioni e usi diversi dell'Intelligenza Artificiale, più consapevoli e più sostenibili: volendo, ad esempio, aumentare la fiducia e ridurre la sfiducia nei confronti dell'IA si potrebbero e si dovrebbero pensare tecnologie e sistemi che risultino più sicuri, rendendoli così anche più affidabili. Per chi si occupa di tecnologie digitali e di intelligenza artificiale, parlare di “*human-centered AI*” significa provare a progettare queste tecnologie considerando la persona come la finalità ultima e quella più alta, per promuovere la quale vanno orientati lo sviluppo e l'applicazione di queste tecnologie.

Le paure emergono dai discorsi degli studenti, le persone più giovani sollecitate a riflettere in merito, che considerano l'impatto e gli effetti più concreti legati alla diffusione e all'applicazione di queste tecnologie, ad esempio per quanto riguarda gli scenari futuri del lavoro. Come pensare allora a conciliare i timori inevitabili e le evoluzioni non prevedibili con l'aspirazione a definire condizioni di vita e modelli di società compatibili con un pieno, più ricco sviluppo umano e sociale? L'etica torna qui ad essere interpellata: i principi dell'equità, dell'eguaglianza e della giustizia rappresentano il tema che ritorna e una sorta di *fil rouge* che accomuna gli interventi di Marta Fasan, Eugenia Lancellotta e Daniel Toscano Lopez.

A partire da una prospettiva di diritto costituzionale, Marta Fasan sottolinea come l'IA, entrando nella vita delle società, induca modificazioni e produca nuovi modelli di assetto sociale, caratterizzati da matrici algoritmiche e digitali che plasmano anche le società. Questi scenari presentano rischi importanti, come quello legato a un controllo e a un possibile “governo” da parte dell'IA sulle persone e le popolazioni. Le nuove tecnologie non sono però neutre nel loro approccio, poiché sono caratterizzate da alcuni *bias* o precomprensioni importanti, che — se applicate su larga scala — possono generare forme di esclusione di gruppi e di discriminazione delle persone. L'equità può rappresentare un principio cui ispirarsi anche nel pensare allo sviluppo di tali tecnologie: a fronte di rischi crescenti di discriminazioni tecnologiche e

di “danni tecnologici” verso le persone e le popolazioni, sono state formulate regole a livello europeo, come l'*AI Act*, e a livello internazionale per provare a ridurre i *bias*, cercando di evitare o limitare conseguenze e trattamenti discriminatori. Gli scenari presenti e futuri interpellano l'etica e il diritto, invitando a riconsiderare la dimensione dell'eguaglianza come criterio ispiratore e orientativo nella progettazione stessa delle tecnologie, pensandole come centrate sulla persona e basate sui diritti fondamentali delle persone e delle comunità. È possibile infatti favorire dinamiche di riconoscimento della diversità anche nella progettazione e quindi nelle applicazioni dell'IA, permettendo così una maggiore inclusione e valorizzando approcci intersezionali.

Per pensare di rispettare e promuovere l'eguaglianza in maniera più specifica e particolare è importante riflettere considerando da vicino gli scenari concreti all'interno dei quali essa va garantita e implementata. Le applicazioni e l'uso dell'Intelligenza Artificiale in alcuni ambiti, tra cui quello della medicina, rappresentano una delle sfide e delle promesse più significative, al presente e in prospettiva: in alcuni settori però le tecnologie digitali non sono ancora state sviluppate in maniera approfondita, dal momento che in essi è più complesso coinvolgere le persone e i pazienti quali “utenti esperti”. Partendo da una panoramica sulle diverse forme di IA, debole, generale e super, Eugenia Lancellotta riflette nel merito dell'IA pensata per la ricerca e la cura delle malattie e dei disagi mentali e rivede la nozione di “ingiustizia epistemica” (Fricker 2007), cercando di comprendere quanto e come essa possa entrare in gioco in quest'ambito particolare. L'ingiustizia epistemica, che viene distinta in testimoniale ed ermeneutica, può risultare amplificata nelle sue diverse dimensioni, così come potrebbe risultare penalizzata la dimensione partecipativa. Fare ricerca sull'IA per quanto riguarda salute e malattia mentale richiede di adottare approcci e sguardi differenti sulle persone che vivono la malattia mentale, favorendo ad esempio la consapevolezza della presenza di dinamiche di ingiustizia epistemica tra chi cura le persone e chi progetta le tecnologie per permettere loro di maturare uno sguardo diverso e una diversa comprensione della condizione specifica e complessa che il disagio e la malattia mentale generano.

Partendo infine da un excursus storico che ripercorre le tappe fondamentali dell'evoluzione dell'IA, Daniel Toscano Lopez sviluppa una

riflessione critica sull'Intelligenza Artificiale, cercando di identificarne la specificità e di capire quale valore essa possa avere in termini etici, sociali e politici. In una prospettiva filosofico-politica l'Intelligenza Artificiale sembra essere caratterizzata da una significativa ambivalenza: accanto alle indiscutibili abilità che essa offre, che fanno pensare ad una sorta di "eccezionalità" che la caratterizza, l'IA genera anche un sentimento di estraneità rispetto a quei fenomeni che vengono percepiti e considerati come più noti e consueti. Analizzando alcuni casi e scenari particolari, Daniel Toscano Lopez propone di osservare le evoluzioni e di pensare agli scenari futuri dell'IA in un'ottica più ampia, che sappia considerare la sua rilevanza biopolitica, accostando tali evoluzioni in un'ottica di responsabilità, di trasparenza, di attenzione pedagogica ed adottando infine, ma non da ultimo, come parametro di giudizio e di valutazione di queste tecnologie la cosiddetta "giustizia globale", orientata a promuovere tramite esse sostenibilità, democratizzazione e partecipazione, così come a prevenirne un uso geopolitico improprio.

A fronte del dibattito in corso sull'IA, che è molto intenso e significativo (Floridi e Cabitza 2021), resta aperto il bisogno di riflettere, confrontarsi e pensare. Come il rapporto con le macchine ci modifica, ad esempio, cosa cambia e potrebbe cambiare nel nostro modo di guardare il mondo, di percepire noi stessi e di entrare in relazione con gli altri? Le nuove tecnologie digitali e l'intelligenza artificiale sembrano in certo qual modo poterci potenziare, perché possono aumentare e amplificare le nostre capacità. Quel che non sappiamo però è come questo aumento di abilità e questo miglioramento di performance possa incidere sul nostro modo di conoscere, di pensare e di percepire, come lo possa modificare e trasformare. Ci possiamo chiedere — ancora — se al potenziamento delle nostre abilità cognitive possa corrispondere anche un miglioramento più complessivo, di noi in quanto esseri umani, un miglioramento della nostra umanità, nei termini ad esempio delle nostre capacità di sentire l'altro, di percepirne le difficoltà e coglierne i punti di forza, così come di intuirne le paure, le debolezze e le vulnerabilità, per esercitare abilità di comprensione e compassione, che forse rappresentano ancora — e ancor di più oggi — la vera sfida per la quale ci potremmo e ci dovremmo impegnare. Alla "Darmouth Conference" sull'Intelligenza Artificiale del 1956, nel corso della quale si era affrontata e

discussa la questione delle macchine intelligenti, ci si era interrogati sulla possibilità di simulare — e non di sostituire — l'intelligenza. Anche se oggi pare che una forma di “sostituzione” stia diventando e sia diventata possibile, resta forse la combinazione — specifica e sapiente — di artificiale e naturale quello che potrà permettere di ottenere forme e modalità nuove di riflettere e agire, anche per quanto riguarda le nostre capacità di interazione, di attenzione e comprensione.

Un'ulteriore preoccupazione emerge quando si entra nel dibattito più sociale e comunitario. Come si può pensare di garantire che il ricorso alle nuove tecnologie possa permettere lo sviluppo e l'esercizio di capacità fondamentali quali la creatività e l'immaginazione, consentendo di mantenere e promuovere la componente della speranza e la dimensione della fiducia, le abilità di interazione e quelle di autoriflessione, delle quali c'è e abbiamo un estremo bisogno, se vogliamo preservare la nostra capacità di essere e rimanere umani? Serve forse ancora, come è stato sottolineato riprendendo un termine e una riflessione di Henri Bergson (Gaillard 2024), dotarsi di un “supplemento d'anima”, che ci permetta di ripensarci e spenderci considerando come promuovere ed esercitare abilità antiche e risalenti secondo modalità nuove, mettendole a servizio di forme d'attenzione per gli esseri e la vita umani, in un rapporto di scambio costante e in una dinamica di ridefinizione e trasformazione continua, che ci e che la caratterizzano.

Entro questo “supplemento d'anima” possiamo ricomprendere alcune tra le facoltà e le abilità che più forse ci caratterizzano, come l'immaginazione, la capacità di intuizione e di interpretazione, la dimensione e lo sguardo poetico con il quale noi esseri umani sappiamo talvolta guardare al mondo, alle persone, alla vita, l'abilità di sentire e cogliere la bellezza dell'altro, degli altri e di quanto ci circonda. Resta aperta anche, infine, la sfida al miglioramento e al potenziamento di quella dimensione che potremmo definire come la nostra capacità d'amare, che richiede una buona dose di intelligenza, di umiltà, di ironia e di ... perdono.

Vediamo, vedremo come e quanto queste nuove tecnologie potranno migliorare e implementare tali capacità non solo di comprendere, ma anche di incontrare e incrociare lo sguardo di un altro essere umano, per coglierne la sua, e la nostra, profonda umanità.

Riferimenti bibliografici

- CORTESI L. (2025) *L'anima umanistica dell'intelligenza artificiale. Tra immaginazione creativa e soggettività simulata*, Aracne, Roma.
- FLORIDI L. e F. CABITZA (2021) *Intelligenza artificiale. L'uso delle nuove macchine*, Bompiani, Firenze.
- FRICKER M. (2007) *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford University Press, Oxford.
- GAILLARD R. (2024) *L'homme augmentée. Futurs de nos cerveaux*, Bernard Grasset, Paris.
- TURING A. (2024) *Macchine calcolatrici e intelligenza*, Einaudi, Torino.