

Tutto il buono dell'IA

Non è necessario che una macchina di IA sia intelligente perché sia efficiente nel portare a termine un compito estremamente complicato. Secondo Luciano Floridi spetta alla filosofia immaginare un buon “matrimonio” tra il verde (dell'ecologia) e il blu (della digitalizzazione), all'insegna dell'etica.

Annamaria Iantaffi



Giornalista pubblicista
e membro dell'Ufficio Stampa e Comunicazione
CISL Scuola Nazionale

È davvero intelligente, l'intelligenza artificiale? Il filosofo Luciano Floridi, già professore ordinario di Filosofia ed etica dell'informazione presso l'Università di Oxford, nominato il 7 settembre Presidente del Comitato Scientifico dell'AAIE (Association of AI Ethicists), non ha dubbi: la risposta è “no”. Non più di un tostapane o di un frigorifero, attualmente. Questa è capace non di produrre “pensando” autonomamente, ma di *agire riproducendo impressionanti comportamenti di calcolo* (e dunque artefatti), i cui effetti, però, possono essere usati a detrimento o in aiuto dell'intera umanità.

Non è una nomina casuale, quella di Floridi a capo dell'AAIE: l'associazione, previa valutazione e convalida dell'*AI Ethicist Certification Program*, vuole stabilire, accanto a una *governance* del digitale, uno standard rigoroso per assicurare che i professionisti dell'etica dell'IA possiedano conoscenze e competenze necessarie per affrontarne le complesse sfide etiche che li (e ci) aspettano(1).

Con una spiccata capacità analitica, non di rado munita di sarcasmo e uno stile chiaro e sequenziale, l'autore procede nel testo (che risale al 2022 ma è ancora estremamente valido nel suo impianto) a una disamina dell'essenza dell'IA e dei suoi sviluppi prevedibili, per poi valutarne i principi etici che attualmente la informano e le pratiche che ne derivano.

Floridi ammonisce il lettore a non considerare l'IA solo un "servizio commerciale" e a non ripetere l'errore che è stato fatto con il web, che è stato normato solo dopo anni dal suo sviluppo di massa, negli anni Novanta. Invita invece a considerare già in atto la quarta rivoluzione, che segue quelle di Copernico, Darwin e Freud, la quale vede gli esseri umani come organismi informazionali, abitanti non unici dell'infosfera(2). L'IA è una "forza potente, una nuova forma dell'agire nell'infosfera che sta già rimodellando le nostre vite, interazioni e ambienti" che va costantemente monitorata e normata, tenendo bene a mente la *direzione* verso cui la si dovrebbe volgere. Ovvero quella kantiana che guarda all'essere umano sempre anche come fine, e non solo come mezzo, per ottenere il bene comune e la protezione dell'ambiente che condividiamo.

A fronte della comparazione(3) dei principali testi internazionali, istituzionali e privati, caratterizzati da alta reputazione e influenza, redatti dal 2017 con valore di indicazione etica, Floridi fa una sintesi di cinque principi che dovrebbero regolare l'agire nell'IA: la *beneficenza*, la *non malevolenza*, l'*autonomia*, la *giustizia* e l'*esplicabilità* (intesa con il duplice valore: epistemologico di intellegibilità, ed etico, di attribuzione della responsabilità). Non sono pochi i comportamenti malevoli, quotidianamente riscontrabili, che Floridi cita nel suo saggio: primi fra tutti lo shopping(4), il *bluwashing*(5), il lobbismo(6) e il *dumping*(7) etici.

Di contro, la disamina di Floridi conduce anche all'individuazione di sette principi di AI4SG(8) (Artificial Intelligence for Social Good), ovvero di esempi di buone pratiche d'uso dell'IA, dichiaratamente volte al bene comune. Tra queste ultime si annoverano ad esempio modelli per prevedere lo shock settico o modelli della teoria dei giochi per prevenire la caccia di frodo. Infine, in ambito scolastico è utile qui ricordare che app di IA generativa sono usate quotidianamente dagli studenti per tradurre, illustrare, sintetizzare, presentare(9), e che l'IA è stata effi-

cacemente usata nel 2015 per anticipare e quindi contrastare la dispersione scolastica su un distretto che comprendeva 200mila studenti statunitensi⁽¹⁰⁾.

La sfida presente e futura dell'AI4SG è dunque applicare l'IA per aiutare la realizzazione del sé, senza svalutare le capacità umane. Il rischio concreto, infatti, è di deprezzare le “vecchie” competenze dei lavoratori e stravolgere il mercato dell'occupazione a livello individuale o societario. Il timore che si profila dietro a un ritmo accelerato di sostituzione di vecchi con nuovi soggetti di competenza è di perdere *know how* ad alta specializzazione ed elevate capacità (ad esempio nella diagnostica sanitaria e nell'aviazione), rendendoci fragili in caso di attacchi avversi o di malfunzionamento dell'IA.

Inoltre, nell'applicazione dell'IA è implicita la possibilità di un enorme potenziamento delle capacità dell'essere umano ma, al contempo, si profila anche il rischio di una sua estrema deresponsabilizzazione, dovuta a quella che Floridi chiama *black-box thinking*, “la mentalità da scatola nera”. Se i processi decisionali adottati con il presidio dell'IA sono percepiti come “oltre la comprensione”, e quindi oltre il controllo umano, si rischia di delegare la responsabilità dell'agire personale. Al contrario, se opportunamente applicati, i sistemi di IA potrebbero amplificare sistemi morali condivisi.

Partito dal divorzio tra l'intelligenza e l'efficienza nel portare a termine un compito (caratteristiche proprie dell'IA), Floridi giunge in chiusura al “matrimonio” tra il verde (dell'ecologia) e il blu (della digitalizzazione), individuando in esso le coordinate auspicabili per il futuro da disegnare. Egli assegna alla filosofia il compito di concettualizzare il design di una società in cui tale connubio sia possibile, e per questo si ripromette di dedicare il prossimo libro alla politica dell'informazione.



(1) Floridi si augura il proliferare di un'etica *soft*, distinta da quella *hard* (che contribuisce a modificare il diritto, ad esempio quello che dovrebbe eliminare il divario di retribuzione tra generi nel mondo), di cui l'etica *soft* condivide l'ambito normativo, ma che “esonda” dai confini giuridici e va al di là della normativa vigente. Egli la definisce “un'etica post *compliance*”.

(2) “Lo spazio semantico costituito dalla totalità dei documenti, degli agenti e delle loro operazioni” in *Infosfera, in Internet & Net Economy*, Vito di Bari (a cura di), Il Sole 24-Ore Libri, 2002

(3) Principi di Asilomar per l’IA (Future of Life Institute, 2017); Dichiarazione di Montréal per l’AI responsabile (Università di Montréal, 2017); Ethically Aligned Design: A vision for Prioritizing Human Wellbeing with Autonomous and Intellingent Systems (IEEE, 2017); Dichiarazione su Intelligenza Artificiale, robotica e sistemi autonomi (EGE, 2018); “Cinque principi generali per un codice di Intelligenza artificiale” in “AI in the UK: ready, willing and able?” (AIUK; House of Lords – Artificial Intelligence Committee, 2018); Principi di partenariato sull’IA (Partnership on AI, 2018).

(4) La tendenza a sfruttare la proliferazione incontrollata dei principi etici per “acquistare” e comporre il quadro di quelli preferiti in base alla propria (personale o aziendale) convenienza.

(5) “Il malcostume di fare affermazioni infondate o fuorvianti a riguardo ... dei valori etici e dei benefici dei processi, prodotti, servizi o altre soluzioni digitali, al fine di apparire più etici dal punto di vista digitale, di quanto non sia effettivamente”.

(6) “Il malcostume di sfruttare l’etica digitale per ritardare, rivedere, sostituire o evitare un’idonea e necessaria regolazione giuridica ... relativa al design, lo sviluppo e l’implementazione di processi prodotti, servizi o altre soluzioni digitali”.

(7) È una espressione coniata nel 2013 dalla Commissione Europea per descrivere l’esportazione in luoghi altri di attività di ricerca su processi, prodotti, servizi digitali contrarie all’etica dei Paesi in cui sono originari, al fine di eluderne le leggi.

(8) Intelligenza Artificiale per il Bene Sociale è un acronimo che indica la disciplina con cui si possono realizzare sistemi informatici seguendo fattori eticamente robusti e pragmaticamente applicabili. Il design di buone pratiche che andrebbero promosse dovrebbe dunque, secondo Floridi, testualmente includere (in modo non esaustivo) 1) la falsificabilità e l’implementazione incrementale; 2) garanzie contro la manipolazione dei predittori; 3) l’intervento contestualizzato in ragione del destinatario; 4) la spiegazione contestualizzata in ragione del destinatario e finalità trasparenti; 5) tutela della privacy e consenso dell’interessato; 6) equità concreta; 7) semantizzazione adatta all’umano.

(9) ChatGPT di OpenAI e Bard di Google sono solo gli esempi più famosi. Ma per la parte iconografica ci sono anche Photopea e UPDF.

(10) Lakkaraju Himabindu, Everal Aguiar, Carl Shan, David Miller, Nasir Bhanpuri, Rayid Ghani, Kecia L. Addison “A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes” in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD, agosto, pp. 1909–1918.

