

ANALISI

Nuovi diritti umani nell'era digitale La tecnica non diventi un destino?

LAURA PALAZZANI

La Dichiarazione di Bletchley, la più recente a livello mondiale, propone una IA sicura, spiegabile, trasparente, inclusiva, giusta ed equa. C'è ampio consenso che l'Intelligenza artificiale debba essere etica, ma servono normative internazionali che riconoscano il primato dell'umanità sul progresso delle nuove macchine. Gli sviluppi accelerati dell'intelligenza artificiale (IA) sono davanti ai nostri occhi. Ne siamo già immersi anche nella quotidianità: nell'ambito del lavoro, dell'economia, della salute, del diritto, ecc. Ne intuiamo le enormi opportunità attuali e future, ma emergono anche preoccupazioni. Quale dunque la parte "buona" della IA su cui investire le nostre speranze e quale la parte da evitare? Stiamo assistendo ad una diffusione dell'attenzione all'etica della IA, ossia alla riflessione sulla distinzione bene/male nell'ambito delle pratiche connesse agli sviluppi della IA. Non è certo nuova l'etica delle tecnologie: si pensi ai decenni di dibattito sulle questioni bioetiche relative alle applicazioni all'inizio e alla fine della vita. La novità è che, se in bioetica emergono spesso, sui temi più delicati, divergenze dell'etica e dunque la difficoltà del bio-diritto, nell'ambito dell'etica della IA si registrano molte convergenze.

Pochi i seguaci delle correnti radicali techno-centriste, favorevoli all'"imperativo tecnologico" che prescrive il "dovere" di applicare sempre la tecnologia in modo incondizionato, anche per superare la condizione umana considerata limitata (si pensi al transumanesimo e al postumanesimo, entrambe espressioni dell'anti-umanesimo): posizioni prevalentemente speculative e minoritarie. Sono ormai molti i documenti internazionali di questi ultimi anni che evidenziano, contro il techno-centrismo, la condivisione dell'umano-centrismo. La prospettiva umano-centrica è la prospettiva di chi parte da due assunti: 1) la tecnologia non è un "destino", ma siamo noi, esseri umani, che costruiamo la tecnologia e possiamo/ dobbiamo dunque definirne i limiti; 2) la costruzione della IA deve essere orientata al riconoscimento della centralità dell'uomo.

Tra i documenti che hanno condiviso questi assunti, si ricorda: The High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, 2019; European Group on Ethics in Science and New Technologies, Statement on Artificial Intelligence, Robotics and 'Autonomous Systems', 2018; OECD, Legal Instruments Artificial Intelligence, 2020; UNESCO, Recommendation on the Ethics of Artificial Intelligence, 2021. Tutti documenti che, seppur con talune distinzioni, registrano una condivisione sui requisiti etici per la regolazione della IA. Ed è interessante verificare che le stesse linee sono tutte presenti anche nella Dichiarazione di Bletchley (1-2 novembre 2023).

Molti gli aggettivi che si trovano più volte ripetuti nel testo per identificare l'etica del ciclo di vita della IA - progettazione, sviluppo, applicazione - con un «controllo umano appropriato»:



Avvenire

«sicura», ossia testata per verificarne la proporzionalità dei benefici attesi rispetto ai rischi potenziali; «spiegabile» e «trasparente», per potere tracciare i passaggi, la logica degli algoritmi e delle correlazioni, limitando l'opacità nei limiti possibili; «inclusiva», non discriminatoria, senza pregiudizi, stigmatizzazioni o esclusioni di accesso di esseri umani per età, genere, etnia, disabilità; «giusta» e «equa», inclusiva per l'accesso alle opportunità, superando il divario digitale, cercando di «non lasciare nessuno indietro»; rispettosa della privacy e della protezione dei dati personali; «sostenibile», rispettosa dell'ambiente per realizzare una transizione digitale compatibile con la transizione ecologica.

Non sfugge, nel testo, uno scivolamento tra l'uso del condizionale "si dovrebbe" (should) e solo in alcuni passaggi l'imperativo "si deve" (must): ma un documento sottoscritto da diversi portatori di interessi, incluse le grandi aziende informatiche, e da diversi Paesi (tra cui Usa e Cina) non poteva che esprimere un accordo generale sui valori condivisi minimi. Ed è già un primo passo importante. La parola più frequentemente utilizzata nel documento è «affidabilità»: l'insistenza è sulla «responsabilità» dei progettisti, dei ricercatori e dei produttori della IA per costruirla «per il bene di tutti» e sull'esigenza di «cooperazione internazionale» per promuovere l'innovazione tecnologica globale.

In questa dichiarazione si avvertono gli influssi del dibattito europeo e l'esigenza di espanderla sul piano globale: in particolare nell'ambito della ricerca dell'impatto potenziale dei sistemi di IA per valutare i rischi incerti e imprevisti, con particolare attenzione ai rischi connessi a usi intenzionali impropri, alla manipolazione dei contenuti o alla generazione di contenuti ingannevoli. I firmatari del documento sono consapevoli che sono di «importanza critica» le riflessioni etiche sulla IA, la prudenza (parola non usata, ma che si legge tra le righe) e la comprensione dei potenziali rischi, dei possibili «danni seri, persino catastrofici», sia deliberati che involontari, affermando «la necessità e l'urgenza di affrontarli». Chiaro anche l'intento ultimo della dichiarazione: quello di promuovere una governance della IA che coinvolga tutti i portatori di interesse, in ambito politico, economico e civile. La Bletchley Declaration è uno dei pochi documenti che esplicitamente menziona l'esigenza di "mettersi intorno ad un tavolo" tra ricercatori, aziende, politici, giuristi, esperti di etica, e coinvolgere nella discussione la società. L'insistenza sulla sicurezza, certamente importante, si scontra con chi ritiene, scientificamente, che algoritmi complessi non potranno mai essere testati esaustivamente: dunque la sicurezza non dovrebbe riguardare solo la progettazione, ma anche la formazione degli sviluppatori ad agire prontamente quando si verificano problemi, sia prima che dopo che i loro modelli sono messi sul mercato.

È dunque indispensabile identificare criteri concordati nella comunità internazionale per una IA sicura, prima e dopo l'impiego, analogamente a quanto si fa per la sperimentazione farmacologica. E, forse anche, che la ricerca sulla IA possa avanzare non solo con finanziamenti privati, ma anche statali, con la garanzia di una maggiore supervisione pubblica.

Manca però nel documento l'impegno a informare ed educare adeguatamente i cittadini per partecipare attivamente e criticamente alla discussione pubblica (la alfabetizzazione alla IA); informazione ed educazione dei ricercatori sia accademici che nelle aziende all'etica del disegno della ricerca, anche

Avvenire

mediante l'introduzione nelle aziende di comitati di etica (sul modello della medicina) per identificare i problemi emergenti nella progettazione e messa sul mercato dei prodotti. Si delineano nella dichiarazione un obiettivo ambizioso: «garantire che i benefici della tecnologia possano essere sfruttati in modo responsabile per il bene e per tutti». Quello che serve ora sono normative internazionali che riconoscano esplicitamente, come è avvenuto in bioetica con la Convenzione sui diritti umani e la biomedicina nel 1997, il primato dell'umano e della umanità sull'avanzamento della tecnologia.

Una normativa basata sull'etica della IA condivisa che riconosca i nuovi diritti umani nell'era digitale: il diritto di accedere alle tecnologie di cui sia dimostrata la sicurezza ed efficacia superando gli scenari di incertezza, mitigando o escludendo i rischi, nella misura possibile; il diritto a non essere 'costretti' indirettamente ad usare le tecnologie, a non essere emarginati o marginalizzati; il diritto ad essere informati ed educati all'uso delle tecnologie ed acquisire una consapevolezza critica per una valutazione autonoma e responsabile; il diritto a partecipare alla governance delle tecnologie, ad essere coinvolti attivamente dalle istituzioni nella definizione delle politiche e nella regolamentazione delle nuove tecnologie. Per ora si è aperto un tavolo, oltre i confini geografici e politici, con un appuntamento al prossimo anno. E forse, nel frattempo, l'approvazione della regolazione europea.

In questo senso, i recenti interventi di Papa Francesco sul tema dell'etica della intelligenza artificiale sono di particolare interesse e auspicano una regolazione etica della IA che riconosca al centro la dignità umana. Nel messaggio per la 58ª Giornata Mondiale delle Comunicazioni Sociali, Intelligenza artificiale e sapienza del cuore: per una comunicazione pienamente umana, Francesco riflette su opportunità e rischi delle nuove tecnologie che stanno «modificando in modo radicale» l'informazione e la comunicazione: avverte dei pericoli delle false notizie e immagini, che alterano la verità; dei rischi della sostituzione dei giornalisti, che devono continuare ad operare sul campo, con capacità critica; della necessità che le regole garantiscano la autenticità, la trasparenza, la responsabilità.

Nel Messaggio per la 57ª Giornata Mondiale della Pace, Intelligenza artificiale e pace, Francesco esprime le preoccupazioni che le nuove tecnologie contribuiscano all'iniquo sviluppo del mercato e del commercio delle armi, e richiama «il cuore» stesso dell'uomo che corre il rischio di diventare sempre più «artificiale» come l'intelligenza. L'auspicio è una regolazione che garantisca la «supervisione umana adeguata, significativa e coerente» e che sappia proibire la «dittatura tecnologica» e il «controllo», che allontanano sempre più la tecnologia dalla pace e dalla giustizia. Ordinario di Filosofia del diritto, Lumsa RIPRODUZIONE RISERVATA Manca nel documento l'impegno a informare ed educare in modo adeguato i cittadini per partecipare attivamente e criticamente alla discussione pubblica circa l'uso degli strumenti.