INTELLIGENZE ARTIFICIALI: ABITARE E GOVERNARE

LUCIO ROMANO*

Il linguaggio suggestivo utilizzato per descrivere le Intelligenze Artificiali (IA) e i loro effetti rileva nel realismo tecnologico. Un vocabolario antropomorfico che non distingue i confini tra ciò che è umano e ciò che descrive i sistemi artificiali. Sono attribuiti termini come intelligenza, comprensione, apprendimento, allucinazione, coscienza: vale a dire che si antropomorfizzano le macchine, creando un ponte semantico ovvero una similitudine tra l'umano e l'artificiale.

Ne conseguono sostanziali questioni filosofiche, giuridiche ed etico-sociali.(1) [48] Mentre si estendono sempre più gli orizzonti delle IA, è fondamentale mantenere consapevolezza critica, promuovendo il bilanciamento tra dimensione umana e dimensione artificiale, evitando reciproche esclusioni, favorendo atteggiamento di fiducia e cautela. Da cui la necessità di abitare e governare le innovazioni tecnologiche per uno sviluppo umano-centrico (*human-centric /human-centred*).

ABITARE

Quando parliamo di intelligenza, coscienza, emozioni o lavoro delle macchine usiamo un linguaggio metaforico che attribuisce qualità di cui sono prive. Il tipo di *λόγος*, di coscienza o di autocoscienza di cui sono dotate è costituito da algoritmi, sequenze di comandi. L'idea di copiare il funzionamento del cervello e della mente umana attraverso l'IA al difuori del contesto delle relazioni, dell'ambiente e della cultura in cui ciascuno è immerso, ha condotto a delle semplificazioni e a degli errori di valutazione fuorvianti. (2)[49]

È essenziale riconoscere i limiti delle metafore linguistiche quando si applicano a entità artificiali, poiché si possono creare ingannevoli analogie con le capacità umane, attribuendo alle macchine le capacità proprie di soggetti-agenti in grado di assumere decisioni in modo consapevole.

---

* Centro Interuniversitario di Ricerca Bioetica (CIRB) - L'articolo è stato redatto in data 9.2.2024

(1)[48] COMITATO NAZIONALE PER LA BIOETICA e COMITATO NAZIONALE PER LE BIOTECNOLOGIE, BIOSICUREZZA E SCIENZE DELLA VITA, *Intelligenza artificiale e medicina: aspetti etici*, Roma 2020; A. PATRONI GRIFFI (a cura di), *Bioetica, diritti e intelligenza artificiale*. Mimesis 2023

(2)[49] R. BODEI, *Dominio e sottomissione*. Il Mulino 2023, p. 300.

Prendiamo ad esempio alcuni termini, tra i tanti di un vocabolario suggestivo, come “comprensione artificiale”, “apprendimento” e “allucinazione”.

La comprensione artificiale (significando la capacità di un sistema di interpretare e riconoscere informazioni in modo simile a un essere umano) andrebbe oltre la mera elaborazione di dati. Vale a dire che la tendenza all'antropomorfizzazione porta a considerare la comprensione artificiale come un equivalente della comprensione umana. Interrogativo: la macchina comprende veramente o è una simulazione sofisticata basata su algoritmi e modelli di apprendimento automatico?

La distinzione tra comprensione umana (intrecciata tra libero arbitrio e coscienza) e comprensione artificiale (correlazione meccanica, inconscia, automatica tra *input* e *output*, ancorata ai dati e agli algoritmi) segna la differenza. Mentre gli algoritmi e i modelli di *machine learning* possono elaborare un’enormità di dati in tempi rapidissimi, manca loro consapevolezza e intenzionalità.

La comprensione artificiale apre a domande ineludibili sulla natura della conoscenza e della coscienza.(3)[50] Si sollevano pressanti interrogativi sulla possibilità di creare macchine con simulazioni sempre più raffinate e sofisticate di processi cognitivi umani. Così per quanto riguarda le implicazioni etiche, giuridiche e politiche quando consideriamo le responsabilità.(4)[51]

Un elemento chiave della comprensione artificiale è un “apprendimento” continuo e automatico. Ecco un altro esempio di linguaggio antropomorfico conferito alle macchine. Il termine apprendimento suggerirebbe un processo simile a quello umano, ma occorre distinguere tra apprendimento artificiale ed esperienza personale propria dell’apprendimento umano. Le macchine apprendono ottimizzando parametri e adattandosi ai dati forniti. Ma questo processo manca dell'esperienza personale data da emozioni, consapevolezza, contesti sociali. Nell’apprendimento artificiale non si rileva la dimensione della soggettività umana “con” significato morale, dato dalla ponderazione consapevole delle circostanze e dalla capacità di discernere gli effetti.

“Allucinazione” è un altro termine dal richiamo chiaramente antropomorfico, mutuato dall’ambito biomedico, un fenomeno attribuito a stati patologici di competenza prevalentemente psichiatrica. Ebbene, il termine allucinazione è ormai trasmigrato nel lessico comune delle IA e usato correntemente quando sistemi di IA generative basate su *Large Language Models* (LLM) – come ChatGPT, Bing, Bard, ecc. – collegano e interpretano erroneamente i dati, producendo informazioni errate. Più dettagliatamente: imprecisioni, errori o vere e proprie invenzioni, per quanto verosimiglianti.

(3)[50] F. FAGGIN, *Irriducibile.* Mondadori 2022
(4)[51] P. SEVERINO (a cura di), *Intelligenza artificiale. Politica, economia, diritto, tecnologia*. LUISS 2022; C. CAPORALE, L. PALAZZANI (a cura di), *Intelligenza artificiale: distingue frequenter. Uno sguardo interdisciplinare*. Edizioni CNR 2023

Appunto, il verosimile senza distinguere il vero dal falso. Secondo l'analisi probabilistica, sulla base di un'enorme mole di dati, i LLM sono addestrati a prevedere la parola successiva ma senza poter comprendere la differenza tra ciò che è corretto e ciò che è sbagliato. E poi c'è il c.d. bias controfattuale, ovvero la tendenza a considerare vera una premessa fattuale errata.

In una recente ricerca della *Stanford University Human-Centered Artificial Intelligence* è stata dimostrata una incidenza dal 69% all'88% di allucinazioni nell'applicazione di LLM in campo giuridico [(5)52,] un settore, questo, molto coinvolto dall'utilizzo di sistemi con IA, come nella c.d. giustizia predittiva. Da un lato i fautori di queste tecnologie ritengono che possano contribuire a ridurre la discrezionalità delle decisioni giudiziarie. Viceversa, con ragioni ben fondate, si rilevano evidenti criticità conseguenti a pregiudizi e discriminazioni dei dati utilizzati e nell'addestramento dei sistemi.[(6)53]

Emblematico è il caso Eric Loomis, cittadino afroamericano degli Stati Uniti, condannato per i reati di ricettazione e resistenza a pubblico ufficiale: pena detentiva di sei anni, quantificata sulla valutazione del rischio di recidiva (*risk assessment*) secondo un sistema di IA sviluppato da un'azienda privata. Particolare non secondario: l'algoritmo usato risultava coperto da brevetto industriale ed era segreto il suo funzionamento [(7)54.]

Le cause delle allucinazioni nell'IA possono essere varie e complesse. Una delle ragioni principali è la sensibilità degli LLM ai dati di addestramento o la presenza di *pattern* che possono fuorviare l'algoritmo. Dati (*input*) ambigui o non rappresentativi possono generare risposte (*output*) erronee (allucinate), che riflettono interpretazioni distorte della realtà con significative implicazioni etiche e di sicurezza. Sono errori a volte così grossolani da essere immediatamente evidenti. L'incessante progresso nell'addestramento di LLM (es.: apprendimento rinforzato con *feedback* umano) comporterà una costante riduzione delle allucinazioni più grossolane per formare modelli generativi più affidabili. Tuttavia, indipendentemente dal fatto che i problemi delle allucinazioni siano adeguatamente affrontati, gli utenti dovrebbero essere consapevoli dello spettro di capacità e limiti delle IA generative. L'uso cauto e il controllo attento dei fatti sono più importanti che mai.

ll problema è epistemologico. La nuova domanda è: qual è il valore della conoscenza comunicata dalle IA? Se la macchina per la comunicazione verosimile si maschera da conoscenza documentata

(5)52 M. DAHL, ET AL. *Hallucinating law: legal mistakes with Large Language Models are pervasive*. Stanford University Human-Centered Artificial Intelligence, Jan 11, 2024
(6)53 U. RUFFOLO (a cura di), *XXVI lezioni di Diritto nell'Intelligenza Artificiale*. Giappichelli 2021.
(7)54 M. BORGOBELLO, *AI e giusto processo, facciamo il punto: le norme, le applicazioni, le sentenze*. Agenda Digitale, 17.1.2024

della realtà, c'è il rischio che possa generare non un'ondata di entusiasmo, ma uno tsunami di illusioni. E poi, di chi è la responsabilità degli effetti conseguenti alle allucinazioni delle IA generative? [(8)55]

Certo, possiamo dire che la ricorrenza nell'uso di un linguaggio antropomorfizzato non è altro che il prevedibile sviluppo dell'originaria attribuzione sostantiva di "intelligenza" ad "artificiale", in cui vanno a sfumarsi i confini tra naturale e artificiale, fino ad essere indistinguibili. Da qui la necessità, ove sia ancora possibile realizzarlo, di riprendere un linguaggio di senso. Che ci dovrebbe far dire, ad esempio, che la macchina, per quanto definibile intelligente, guarda (con gli occhi della telecamera) ma non vede; ascolta (con le orecchie dei microfoni) ma non sente. È automatica nell'apprendimento (*machine learning, deep learning, reti neurali artificiali, …*) ma non è autonoma. Perché il vedere e il sentire, a differenza del guardare e dell'ascoltare, coinvolgono un'assunzione di consapevolezza o di coscienza che trasformano chi guarda e ascolta in colui che comprende.

Così per quanto riguarda il rapporto tra cervello e pensiero che sarebbe ridotto a una visione fisicalista o funzionalista. Il tema è molto vasto e richiederebbe più ampie riflessioni e argomentazioni. Comunque, "ora accade che, molto spesso, coloro che si propongono di imitare non già le funzioni, ma direttamente la struttura dell'attività pensante umana, identifichino tale attività con il funzionamento del *cervello* umano, di cui il pensiero non sarebbe altro che un prodotto. Se si arrivasse a costruire un cervello artificiale, questo saprebbe quindi pensare. Alla base di questa prospettiva sta però un presupposto gratuito: appunto l'identificazione del pensiero con un *prodotto immateriale del cervello*. Anche in questo caso è rilevabile un errore di logica, ossia l'aver scambiato una condizione *necessaria* per una che sia anche *sufficiente*: è del tutto plausibile che non si dia pensiero senza cervello (per lo meno fin che si tratta di pensiero umano), ma non è detto che basti un cervello per avere il pensiero. Ma c'è anche di più: volendo essere rigorosi, la costruzione di un cervello artificiale non potrà mai consistere nella *esatta riproduzione* di un cervello umano e tutti i progressi della bionica riusciranno al massimo a produrre un artefatto che *funziona* in modo molto simile al cervello naturale"[(9)56.] I dati (*input*) e gli algoritmi, invece, elaborano delle risposte (*output*) senza una spiegazione di senso. Generano contenuti simili a quelli umani, come nel caso di IA generative, sulla base di un mero metodo probabilistico.

È il congedo dagli scenari umanistici? Interrogativo che pone Umberto Galimberti. "E' la logica lineare tipica delle macchine che, prescindendo dall'individuo come soggetto e da tutti i suoi aspetti di vita individuale, lo sostituisce con un profilo dei suoi comportamenti, dei suoi interessi e dei suoi atti.

(8)[55] L. DE BIASE, *Le illusioni dell'intelligenza artificiale*. Nòva, Il Sole 24 Ore, 19.2.2023
(9)[56] E AGAZZI, *Dimostrare l'esistenza dell'uomo*. Mimesis 2023, p.181.

E tutto ciò senza una vera consapevolezza, anzi con la persuasione che la decisione sia ancora nelle loro mani, perché in fondo sono loro ad aver deciso di farsi guidare da altri. Avvertire la distanza che esiste tra il mondo-della-vita e il mondo digitale non significa delegittimare l'intelligenza artificiale, ma non smarrire la singolarità e la specificità che caratterizza ogni vita umana e in generale tutto ciò che di essa non è traducibile in un algoritmo." (10)[57]

L'attribuzione ai sistemi di IA di proprietà umane attraverso l'uso di un vocabolario antropomorfico porta inevitabilmente a una visione riduzionista o materialista. L'uso dell'aggettivo intelligente alla macchina provoca la domanda filosofica: se la macchina pensa come l'uomo, l'uomo pensa come la macchina? Sembrano derivarne due conseguenze possibili: o si ammette che anche la (nuova) macchina possa compiere operazioni in quanto dotata di un potere immateriale, come il pensiero; oppure si deve ammettere che, se una macchina materiale sa fare cose intelligenti, allora anche nel caso dell'uomo non è necessario ammettere un principio immateriale per fare tali cose. In entrambi i casi si avrebbe una riduzione dell'uomo a un livello strettamente *naturalistico* (se ne perde la specificità rispetto al mondo naturale). Nel secondo caso si avrebbe anche una connotazione esplicitamente *materialista*. (11)[58]

In sintesi, potremmo dire che l'assegnazione di categorie lessicali propriamente umane alle macchine significa una delega suggestiva, o inquietante se si vuole. Non solo semantica ma essenzialmente concettuale, che accompagnerà inevitabilmente l'evoluzione dei sistemi di IA. Insomma, un nuovo lessico antropomorfico con una nuova narrazione?

GOVERNARE

Nella consapevolezza dei limiti in grado di generare rapidamente quantità diverse e rilevanti di disinformazione convincente, si accredita l'urgente necessità di un governo ovvero di una regolamentazione. Accanto ai progressi, i rischi emergenti stanno diventando sempre più evidenti, in una molteplicità di settori. Trasparenza, sorveglianza e regolamentazione possono fungere da preziosi riferimenti per la gestione dei rischi, a fronte, poi, della progressione degli investimenti globali sull'IA: si prevedono per il 2024 investimenti per circa 300 milioni di dollari e oltre 420 per il 2025. Per il 2030, ben 1900 miliardi di dollari.

Si delineano, così, nuove strutture di potere e si configura un nuovo decisore sociale. Un "capitalismo della sorveglianza".(12) [59] Da cui la necessità di un codice etico ovvero di una

(10)[57] U. GALIMBERTI, *L'etica del viandante*. Feltrinelli 2023, p.27
(11)[58] E. AGAZZI, *Dimostrare* … p.178
(12)[59] S. ZUBOFF, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*. LUISS University Press 2019

regolamentazione. Un percorso che dovrebbe avere tre obiettivi sostanziali: consapevolezza nell'abitare le innovazioni tecnologiche; interdisciplinarità delle competenze; regolamentazioni, secondo principi condivisi di etica sociale, dell'innovazione tecnologica nell'ottica dello sviluppo umano-centrico.[(13)60]

Obiettivi, questi, certo non facili da raggiungere, tuttavia ineludibili. Da un lato vi è il riduzionismo antropologico del paradigma tecnocratico, dominio degli algoritmi o algocrazia.[(14)61] Dall'altro lo sviluppo umano-centrico (*human-centric /human-centred*), che tiene conto delle innovazioni tecnologiche ma che sa coniugarle secondo un'etica degli algoritmi ovvero "algoretica".[(15)62] In particolare la Pontificia Accademia per la Vita con la *Rome Call for AI Ethics* ha indicato sei principi etici di riferimento: trasparenza (i sistemi di IA devono essere comprensibili a tutti), inclusione (nessuna discriminazione perché ogni essere umano ha pari dignità), responsabilità (ci deve essere sempre qualcuno che si assume la responsabilità di ciò che fa una macchina), imparzialità (i sistemi di IA non devono seguire o creare pregiudizi), affidabilità, sicurezza e privacy.[(16)63]

Potremmo delineare, quindi, un orizzonte di un umanesimo digitale che "non trasforma l'essere umano in una macchina e non interpreta le macchine come esseri umani. Che riconosce la peculiarità dell'essere umano e delle sue capacità, servendosi delle tecnologie digitali per ampliarle, non per restringerle." [(17)64]

Comunque, sono molteplici i principi etici proposti. Come richiamato da Luciano Floridi, "l'enorme volume di principi proposti – più di 160 nel 2020, secondo l'*AI Ethics Guidelines Global Inventory* (la rassegna globale delle linee guida etiche per l'IA) di *Algorithm Watch* – rischia di diventare soverchiante e fuorviante, sollevando due potenziali problemi. O i vari insiemi di principi etici per l'IA sono simili, portando a inutili ripetizioni e ridondanze, oppure, se differiscono in modo significativo, sono suscettibili di ingenerare confusione e ambiguità. Il peggior risultato sarebbe quello di creare un mercato di principi in cui le parti interessate potrebbero essere tentate di acquistare quelli più allettanti."[(18)65]

Allora come risolvere il problema della proliferazione dei principi? Sempre Floridi, sulla base di un'analisi comparativa, identifica "un quadro generale costituito da cinque principi fondamentali per l'IA etica. Prendendo atto dei limiti e valutando le implicazioni di questo quadro etico per gli sforzi

(13)[60] C. CASONATO, *L'intelligenza artificiale e il diritto pubblico comparato ed europeo*, DPCE OnLine 2022.
(14)[61] FRANCESCO, *Intelligenza artificiale e pace. Messaggio per la LVII Giornata Mondiale della Pace*. Città del Vaticano, 1.1.2024.
(15)[62] P. BENANTI, *Human in the loop. Decisioni umane e intelligenze artificiali*. Mondadori 2022, p. 133
(16)[63] PONTIFICAL ACADEMY FOR LIFE -RENAISSANCE FOUNDATION, *Rome Call for AI Ethics*. Roma, February 28th 2020
(17)[64] J. NIDA-RUMELIN, N. WEIDENFELD, *Umanesimo digitale*. Franco Angeli 2019, p.8; L. ROMANO, *Enhancement cognitivo e nuovo umanesimo digitale,* in A. PATRONI GRIFFI (a cura di), *Bioetica, diritti e intelligenza artificiale*. Mimesis 2023, p.137
(18)[65] L. FLORIDI, *Etica dell'intelligenza artificiale*. Raffaello Cortina Editore 2022, p.92

futuri volti a creare leggi, regole, standard e le migliori pratiche (*best practices*) per l'IA etica in un'ampia varietà di contesti." I cinque principi individuati sono: beneficenza, non maleficenza, giustizia, autonomia ed esplicabilità. I primi quattro ordinano, classificano e raggruppano norme etiche che, come da consolidate riflessioni bioetiche, necessitano di contenuto, specificità e gerarchizzazione. Il quinto e nuovo principio di esplicabilità è da intendere "sia in senso epistemologico di *intelligibilità* – come risposta alla domanda: come funziona? – sia in quello etico di *responsabilità* (*accountability*) come risposta alla domanda: chi è responsabile del modo in cui funziona?"

Varie istituzioni, nazionali e internazionali, stanno sviluppando percorsi con la finalità di regolamentazioni condivise, pur nella consapevolezza della complessità del tema e dei rilevantissimi interessi coinvolti.(19) [66] I procedimenti di regolamentazione in corso significano processi di democratizzazione, evitando, così, che la c.d. "scatola nera" (*black box*), ovvero l'inesplicabilità di modelli di IA – rispetto a modelli di IA interpretabili e spiegabili – possa rappresentare un alibi per l'accettazione passiva di qualsiasi procedura o risultato. Accantonando, erroneamente, l'esplicabilità ovvero i chiarimenti sulle cause delle decisioni (*output*), i meccanismi di elaborazione e interpretazione dei dati, la valutazione dell'appropriatezza o inadeguatezza dei risultati prodotti. La strada da seguire coniuga dimensione umana e dimensione artificiale, favorendo l'incontro tra la fiducia e la cautela.

Ecco la necessità di governare con regolamentazioni, abitando le innovazioni tecnologiche nella consapevolezza e cooperando nell'interdisciplinarità delle competenze.

Riporto, in particolare, alcuni dei più recenti documenti pubblicati. In particolare, il *Decalogo* del Garante per la Protezione dei Dati Personali (GPDP); l'*Artificial Intelligence Act* dell'Unione europea (EU AI Act); il Rapporto intermedio *Governing AI for Humanity* dell'organo consultivo del Segretario generale delle Nazioni Unite su rischi, opportunità e governance internazionale dell'IA; la Guida su *Ethics and governance of artificial intelligence for health. Guidance on large multi-modal models* della World Health Organization.

Il GPDP ha emanato il Decalogo per la realizzazione di Servizi Sanitari Nazionali attraverso sistemi di IA. Sono richiamati tre principi: conoscibilità, non esclusività della decisione algoritmica e non discriminazione algoritmica. Per quanto riguarda la conoscibilità, l'interessato ha il diritto di conoscere l'esistenza di processi decisionali basati su trattamenti automatizzati e, in tal caso, di ricevere informazioni significative sulla logica utilizzata, sì da poterla comprendere. Per quanto riguarda la non esclusività della decisione algoritmica, nel processo decisionale deve esistere un intervento umano

(19)[66] G. FINOCCHIARO, *Intelligenza artificiale. Quali regole?* Il Mulino 2024

capace di controllare, validare ovvero smentire la decisione automatica (c.d. *human in the loop*). *Last but not least*, la non discriminazione algoritmica: il titolare del trattamento utilizzi sistemi di IA affidabili che riducano le opacità, gli errori dovuti a cause tecnologiche e/o umane, verificando periodicamente l'efficacia anche alla luce della rapida evoluzione delle tecnologie impiegate, delle procedure matematiche o statistiche appropriate per la profilazione, mettendo in atto misure tecniche e organizzative adeguate.(20)[67]

L'Unione europea con l'*Artificial Intelligence Act* (EU AI Act) si è prefissa l'obiettivo di assicurare che i cittadini europei possano beneficiare delle nuove tecnologie in conformità ai valori, ai diritti fondamentali e ai principi dell'Unione.(21) [68] L'impostazione nella classificazione dei rischi prevede quattro categorie: minimi, limitati, alti, inaccettabili. Tanto maggiore sarà il rischio, quanto più elevate saranno anche le responsabilità, anche per chi sviluppa o adopera sistemi di IA.

In sintesi, l'accordo vieta sistemi di IA che presentano un livello di rischio inaccettabile per la sicurezza delle persone, come quelli utilizzati per il punteggio sociale (classificare le persone in base al loro comportamento sociale o alle loro caratteristiche personali). Divieti, poi, sussistono sugli usi intrusivi e discriminatori dell'IA. Tra questi, l'uso di sistemi di identificazione biometrica remota in tempo reale in spazi accessibili al pubblico; sistemi di categorizzazione biometrica basati su caratteristiche sensibili (ad esempio genere, etnia, cittadinanza, religione, orientamento politico); sistemi di polizia predittiva (basati su profilazione, ubicazione o comportamenti criminali passati); sistemi di riconoscimento delle emozioni; estrazione non mirata di dati biometrici da Internet o da filmati di telecamere a circuito chiuso per creare database di riconoscimento facciale (in violazione dei diritti umani e del diritto alla privacy).

L'ONU, a sua volta, tramite l'organo consultivo sull'IA del Segretario generale delle Nazioni Unite, ha lanciato l'*Interim Report Governing AI for Humanity*(22) [69] titolo significativo ed evocativo. Elemento centrale è la proposta di governance dell'IA, individuando i principi di riferimento. Inclusività: tutti i cittadini dovrebbero essere in grado di accedere e utilizzare in modo significativo gli strumenti di IA. Interesse pubblico: la governance dovrebbe andare oltre il principio di non nuocere e definire un quadro di responsabilità più ampio per le aziende che costruiscono, distribuiscono e controllano l'IA, nonché per gli utenti. Centralità della governance dei dati: la governance dell'IA non

(20)[67] GPDP, *Decalogo per la realizzazione di servizi sanitari nazionali attraverso sistemi di Intelligenza Artificiale*. Roma 2023
(21)[68] Alla data di redazione del presente Articolo, non si è ancora concluso l'iter formale del Regolamento.
(22)[69] UNITED NATIONS, *Interim Report: Governing AI for Humanity*. December 2023.

può essere separata dalla governance dei dati. Universalità, interconnessione e multilateralità: la governance dell'IA dovrebbe dare priorità al consenso universale da parte dei Paesi e delle parti interessate nonché dovrebbe sfruttare le istituzioni esistenti attraverso un approccio di rete. Infine, il Diritto internazionale: la governance dell'IA deve essere ancorata alla Carta delle Nazioni Unite, al diritto internazionale sui diritti umani e agli obiettivi di sviluppo sostenibile. L'organo consultivo sull'IA ha promosso per i prossimi mesi consultazioni aperte con tutte le parti interessate – individui, gruppi e organizzazioni – così da offrire *feedback* prima della stesura definitiva del *Report*.

Recentemente l'OMS ha pubblicato il documento *Ethics and governance of artificial intelligence for health. Guidance on large multi-modal models*. [(23)70] Con la Guida sull'etica e sulla governance dell'IA per la salute, si identificano sia i potenziali benefici che i potenziali rischi dell'uso dell'IA nell'assistenza sanitaria. Sono stati definiti sei principi da prendere in considerazione nelle politiche e nelle pratiche di governi, sviluppatori e fornitori che utilizzano l'IA. I principi sono: proteggere l'autonomia; promuovere il benessere umano, la sicurezza umana e l'interesse pubblico; garantire trasparenza, spiegabilità e intelligibilità; promuovere la responsabilità e l'affidabilità; garantire inclusività ed equità; promuovere un'IA sostenibile. L'OMS ha pubblicato questa Guida per assistere gli Stati membri nella mappatura dei benefici e delle sfide associati all'uso dell'Intelligenza Artificiale generativa che si basa sui LLM. La Guida include raccomandazioni per la governance, all'interno delle aziende, da parte dei governi e attraverso la collaborazione internazionale.

In definitiva, il comune denominatore, costituito da consapevolezza-interdisciplinarità-regolamentazione, rappresenta la prospettiva per un responsabile sviluppo umano alla luce delle rapidissime innovazioni tecnologiche. Non è plausibile, come postulato da tecnofobi e apocalittici secondo un improponibile neo-luddismo, opporsi alle innovazioni tecnologiche. Né è ragionevole la posizione di tecnofili e integrati che ritengono qualsiasi forma di innovazione tecnologica come rappresentazione, di per sé immediatamente condivisibile, di sviluppo umano. Se è vero, come è vero, che "*when you are finished changing, you are finished*" (Benjamin Franklin), ciò non significa accantonare l'umano-centrismo a favore del tecnocentrismo. L'orizzonte è quello di un'antropologia integrale anche per le nuove opportunità.

(23)70 WORLD HEALTH ORGANIZATION, *Ethics and governance of artificial intelligence for health*. Jan 19, 2024