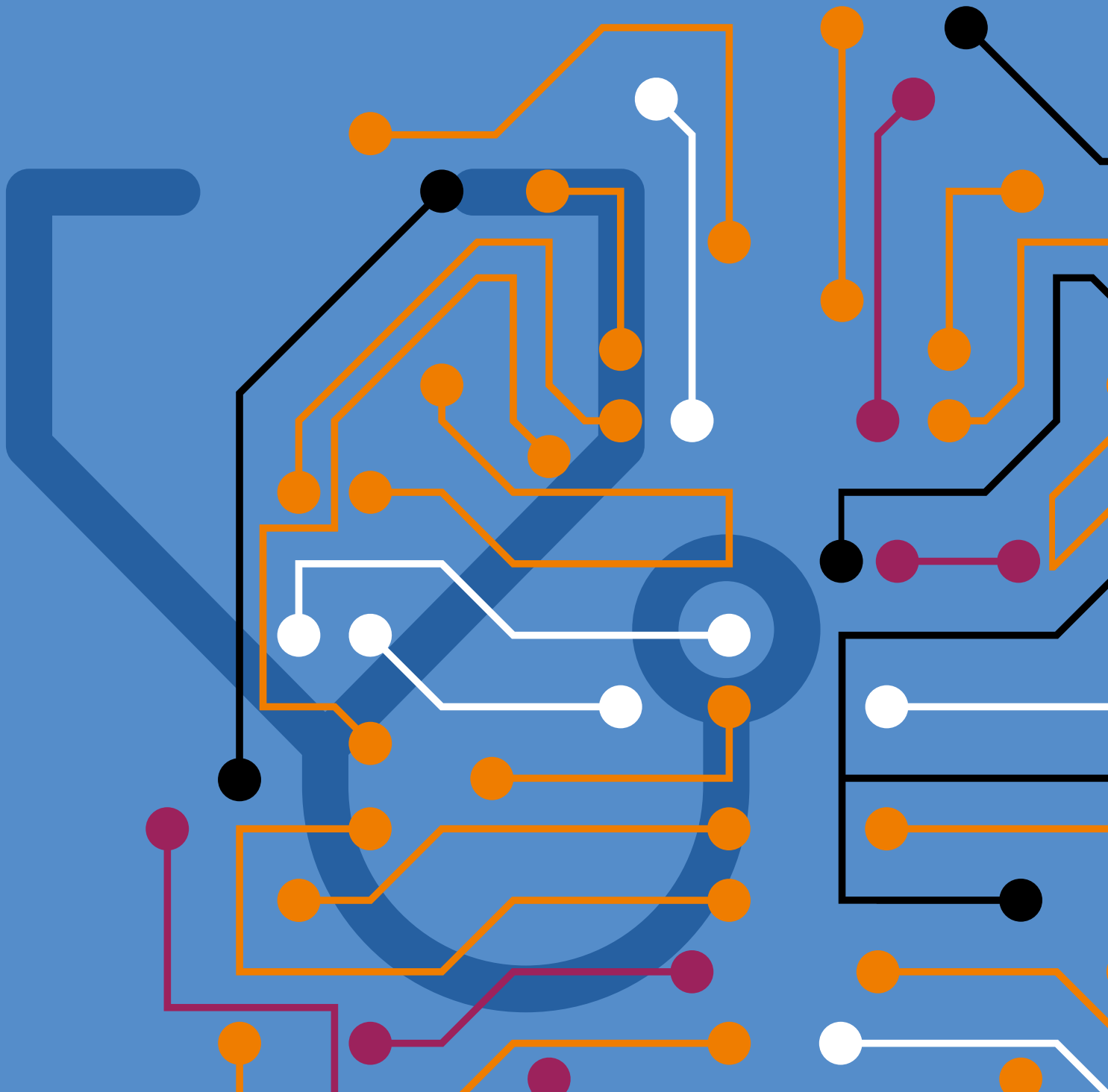
SIIAM
Società Italiana Intelligenza Artificiale in Medicina

SOCIETÀ BENEFIT
ZADIG

Etica e governance dell'intelligenza artificiale per la salute

Linee guida per i modelli multimodali di grandi dimensioni (LMM)



Società Italiana Intelligenza Artificiale in Medicina (SIIAM)

Coordinamento: Francesco Andrea Causio e Angelo Talio

Autori: Alessandro Belpiede, Antonio Baldassarre, Filippo Bonaldi, Claudia Cosma, Luigi De Angelis, Marcello Di Pumpo, Giacomo Diedenhofen, Enrica Frasson, Francesca Giovanetti, Davide Golinelli, Vittorio Greco, Giuseppa Granvillano, Guido Marchi, Alessandro Mazzotta, Andrea Nappi, Cosimo Savoia, Nicolò Scarsi, Michele Tadiello, Leonardo Tariciotti

Zadig srl Società Benefit

Coordinamento: Pietro Dri

Revisione ed editing: Sergio Cima, Maria Rosa Valetto

Traduzione e adattamento della linea guida OMS: “Ethics and governance of artificial intelligence for health. Guidance on large multi-modal models. Geneva: World Health Organization; 2024. Licence: CC BY-NC-SA 3.0 IGO”

La traduzione non è stata fatta dall’OMS che quindi non è responsabile per il contenuto o l’accuratezza della traduzione. Solo l’edizione originale inglese è l’edizione autentica e ufficiale.

Editore: Zadig srl Società Benefit, 2024

ISBN 9788831306379

Indice

Introduzione all'edizione italiana 5

Abbreviazioni 13

Executive Summary 14

1. Introduzione 23

1.1 Rilevanza dei modelli multimodali di grandi dimensioni (LMM) 25

1.2 Linee guida OMS sull'etica e la governance dell'IA in sanità 27

I. Applicazioni, sfide e rischi degli LMM 28

2. Applicazioni e sfide connesse all'utilizzo degli LMM in ambito sanitario 29

2.1 Diagnosi e assistenza clinica 29

2.2 Applicazioni centrate sul paziente 33

2.3 Compiti amministrativi e funzioni di trascrizione 36

2.4 Formazione medica e infermieristica 38

2.5 Ricerca medica e scientifica e sviluppo di nuovi farmaci 38

3. Rischi per i sistemi sanitari e la società e preoccupazioni etiche sull'uso degli LMM 41

3.1 Sistemi sanitari 41

3.2 Conformità ai requisiti regolatori e legali 44

3.3 Rischi e preoccupazioni per la società 45

II. Etica e governance degli LMM nella sanità e nella medicina 51

4. Progettazione e sviluppo di modelli di base per finalità generali 54

4.1 Rischi che devono essere gestiti durante lo sviluppo di modelli di base per finalità generali 54

4.2 Misure che gli sviluppatori possono adottare per gestire i rischi con i modelli di base per finalità generali 55

4.3 Leggi, politiche e investimenti del settore pubblico 59

4.4 LMM open source 62

5. Disposizioni relative ai modelli di base per finalità generali 65

5.1 Rischi che devono essere affrontati quando si fornisce un servizio o un'applicazione sanitaria con un modello di base per finalità generali 65

5.2 Misure che i governi possono introdurre per gestire i rischi e i principi etici da osservare 66

6. Distribuzione di modelli di base per finalità generali 73

6.1 Rischi da gestire nella distribuzione di un servizio o applicazione sanitaria con un modello di base per finalità generali 73

6.2 Responsabilità continua di sviluppatori e fornitori durante la distribuzione 73

6.3 Responsabilità dei distributori 74

6.4 Programmi governativi e pratiche 75

7. Responsabilità per gli LMM 78

8. Governance internazionale degli LMM 80

Riferimenti bibliografici 82

Allegato – Metodi 91

Introduzione all'edizione italiana

Il documento dell'Organizzazione Mondiale della Sanità (OMS) "Ethics and governance of artificial intelligence for health: guidance on large multi-modal models", tradotto a partire da pagina 14 affronta le sfide etiche e di governance associate all'uso dei Large Multi-Modal Models (LMM, modelli multimodali di grandi dimensioni) in ambito sanitario. Questo documento nasce dalla necessità di aggiornare i contenuti del precedente "Ethics and governance of artificial intelligence for health" pubblicato nel 2021 dall'OMS. L'aggiornamento è stato dettato non tanto da nuovi utilizzi dell'intelligenza artificiale (IA) in sanità, quanto dall'emergere di una nuova tecnologia, quella degli LMM, la cui accessibilità rappresenta una novità assoluta, che apre nuovi scenari in termini di benefici e rischi.

La Società Italiana di Intelligenza Artificiale in Medicina (SIIAM) riconosce l'importanza di questo documento nel guidare l'adozione responsabile e sostenibile di queste tecnologie nei sistemi sanitari di tutto il mondo. Affinché queste linee guida possano essere utili ai singoli Stati è necessario interpretarle sulla base delle specificità del contesto nel quale si intendono applicare. Da questa considerazione nasce il lavoro di traduzione e commento delle linee guida OMS portato avanti dalla SIIAM, in collaborazione con l'editore medico-scientifico Zadig, con l'obiettivo di calare le indicazioni dell'OMS nel contesto italiano.

Applicazioni, sfide e rischi degli LMM

La letteratura scientifica propone una varietà di possibili implementazioni per gli LMM in vari campi dell'assistenza sanitaria e della ricerca. Un esempio è rappresentato dalla riduzione del carico di lavoro amministrativo per i medici e gli infermieri. Secondo uno studio pubblicato nel 2016, l'attività di documentazione amministrativa può occupare dal 25% al 50% del tempo di un medico e il 20% del tempo di un infermiere[1]. L'automazione di alcune attività, come la compilazione delle cartelle cliniche elettroniche, la fatturazione e la programmazione degli appuntamenti potrebbe liberare tempo prezioso per gli operatori sanitari, consentendo loro di dedicarsi maggiormente all'assistenza diretta dei pazienti.
Inoltre, gli LMM possono contribuire a migliorare la comunicazione medico-paziente, semplificando il gergo medico e rendendo le informazioni più accessibili ai pazienti. L'uso dell'intelligenza artificiale, in particolare dei Large Language Models (LLM, modelli linguistici di grandi dimensioni), per semplificare il processo di consenso informato in Italia po-

[1] Sinsky C, Colligan L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Ann Intern Med 2016;165:753-60.
Lin S, Mahoney M, et al. Ten ways artificial intelligence will transform primary care. J Gen Intern Med 2019;34:1626-30.

trebbe essere vantaggioso, ma deve essere implementato con cautela e nel rispetto delle disposizioni della legge 219/2017.
Gli LMM potrebbero essere utilizzati per generare spiegazioni semplificate delle condizioni mediche, delle procedure e dei trattamenti, rendendo le informazioni più accessibili e comprensibili per i pazienti e per creare moduli di consenso informato personalizzati, basati sulle specifiche condizioni di salute e sulle esigenze individuali del paziente. Tuttavia, è fondamentale sottolineare che l'uso degli LMM dovrebbe essere visto come un supporto e non come una sostituzione alla comunicazione diretta tra medico e paziente, poiché la legge 219/2017 evidenzia l'importanza della relazione di cura e di fiducia basata sul consenso informato. Inoltre, l'uso degli LMM nell'ambito del consenso informato deve essere sottoposto a un'attenta supervisione da parte dei medici, che rimangono responsabili di garantire che i pazienti ricevano informazioni accurate, complete e comprensibili, ed eventuali errori o imprecisioni generate dagli LMM potrebbero avere conseguenze legali per i medici.

Sicuramente gli LMM sono in grado di rispondere correttamente a domande a risposta multipla nell'ambito medico, in particolare sono state valutate le loro performance nei test somministrati a studenti o giovani professionisti per l'abilitazione alla professione medica[2]. Vi sono alcuni esempi di come poter utilizzare questi strumenti a supporto delle attività didattiche anche in Italia[3], ma non è semplice passare da singoli casi di utilizzo a un'integrazione organica di strumenti di IA a supporto delle attività didattiche all'interno dei curricula delle università. Il tema dell'alfabetizzazione digitale è di particolare interesse e in tal senso è opportuno promuovere attività di formazione continua che vertano su questi argomenti.

L'uso degli LMM presenta molte opportunità quando implementati come strumenti di supporto alla decisione clinica (Clinical Decision Support Systems, CDSS) ma anche rischi importanti, molti dei quali condivisi con i modelli già utilizzati nella pratica clinica. Gli LMM possono produrre risultati inaccurati a causa di dati di input errati o limitazioni del modello. I CDSS basati sull'apprendimento automatico sono simili agli strumenti di supporto clinico sviluppati utilizzando modelli statistici classici e, come tali, hanno limitazioni simili. L'IA può altresì replicare o esacerbare i bias preesistenti nei dati di addestramento, influenzando negativamente le decisioni sanitarie. Se un LMM viene addestrato utilizzando dati che non corrispondono ai dati che incontrerà durante la distribuzione, le sue prestazioni potrebbero essere inferiori al previsto. Per ovviare a ciò occorre un'attenta valutazione degli LMM utilizzando i nuovi dati provenienti dalle prestazioni durante la distribuzione, compresi i campioni di dati che dovrebbero "ingannare" il modello, come quelli con diversi dati demografici della popolazione, condizioni difficili o input di cattiva qualità.

Ridurre i bias legati ai modelli medici attuali e ad alcune caratteristiche dei pazienti, come

---

[2] Kung T, Cheatham M, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digital Health 2023;DOI:10.1371/journal.pdig.0000198.

[3] Baglivo F, De Angelis L, et al. Exploring the possible use of AI chatbots in public health education: feasibility study. JMIR Med Educ 2023;9:e51421.

il genere o la classe sociale, è fondamentale per affrontare i problemi di equità esistenti anche in un Sistema Sanitario Nazionale universalistico come quello italiano.
È inoltre importante considerare come gli LMM possano incentivare il cosiddetto "bias dell'automazione" tra gli operatori sanitari e i pazienti, portando a un eccessivo affidamento su queste tecnologie. Ciò può portare a trascurare errori che normalmente verrebbero rilevati o a delegare in modo inappropriato decisioni complesse agli algoritmi. Inoltre, anche gli LMM sono soggetti a rischi di sicurezza informatica. Tali vulnerabilità possono compromettere la sicurezza dei dati dei pazienti o l'affidabilità degli algoritmi stessi, influenzando negativamente l'erogazione dei servizi sanitari.

Nella ricerca medica, gli LMM sono ormai correntemente utilizzati come supporto per le attività di ricerca scientifica, in particolare nella fase di stesura e revisione degli articoli divulgativi e scientifici. Un ulteriore utilizzo di grande rilievo è nel supporto all'analisi dei dati, visto che questi modelli sono in grado di fornire assistenza nella scrittura di codice.
Sebbene molti gruppi editoriali indichino la necessità di dichiarare l'utilizzo che è stato fatto degli LMM in fase di sottomissione di un articolo a una rivista, non vi sono regole chiare e univoche su quale sia il perimetro di utilizzo ritenuto legittimo. Recentemente vi sono stati esempi di come porzioni di testo e immagini generate con gli LMM possano sfuggire al controllo che dovrebbe essere esercitato da autori, revisori ed editori, portando alla pubblicazione su riviste indicizzate di contenuti generati da IA senza supervisione umana e di scarsa qualità scientifica[4]. Questo da un lato fa emergere le evidenti criticità del sistema di pubblicazione scientifica e dall'altro sottolinea i rischi di un'infodemia potenziata dagli LMM[5].

Considerazioni di natura etica e normativa sugli LMM nella ricerca e nell'ambito medico

Gli LMM vengono implementati più velocemente della nostra capacità di comprenderne appieno le potenzialità e le fragilità. La governance degli LMM deve tenere il passo con il loro rapido sviluppo e con il loro crescente utilizzo, senza privilegiare né i governi che cercano un vantaggio tecnologico né le aziende che perseguono un guadagno commerciale. I principi etici e gli obblighi in materia di diritti umani devono essere al centro di una governance appropriata, che comprenda sia le procedure e le pratiche che potrebbero essere introdotte dalle aziende, sia le leggi e le politiche emanate dai governi.

Il presente documento, nella sua versione originale, ha un approccio molto generico nel tracciare le linee di collaborazione a livello internazionale in ambito di IA. La proposta di coinvolgere una vasta gamma di attori, incluse istituzioni pubbliche, organizzazioni internazionali e la società civile, per definire un quadro normativo che sia equo, inclusivo e rispettoso dei principi legali ed etici fondamentali, rivela un approccio olistico e lungimirante. La creazione di un'agenzia di ricerca pubblica finanziata da diversi governi per guidare lo sviluppo delle forme avanzate di IA e garantire la traspa-

[4] https://scienceintegritydigest.com/2024/02/15/the-rat-with-the-big-balls-and-enormous-penis-how-frontiers-published-a-paper-with-botched-ai-generated-images/

[5] De Angelis L, Baglivo F, et al. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. Front Public Health 2023;DOI:10.3389/fpubh.2023.1166120

renza e la condivisione delle conoscenze rappresenta un passo significativo verso una governance responsabile e partecipativa nel settore dell'IA.
L'analisi critica della proposta di adottare un approccio simile a quello utilizzato per le armi nucleari nel regolamentare l'IA evidenzia la complessità e l'urgenza della sfida che comporta la gestione delle tecnologie avanzate e potenzialmente distruttive. Questo richiede un impegno globale e multilaterale per garantire che l'IA sia sviluppata e utilizzata in modo etico, sicuro e in linea con i valori umani fondamentali.

La preoccupazione che emerge nel documento rispetto all'aderenza alle normative vigenti da parte degli LMM è quanto mai pertinente in Italia, dove il 31 marzo 2023 l'accesso a ChatGPT è stato temporaneamente bloccato in seguito all'istruttoria aperta dal Garante per la protezione dei dati personali[6]. Il Garante contestava ad OpenAI l'assenza di una base giuridica che giustifichi la raccolta e la conservazione massiccia di dati personali allo scopo di addestrare gli algoritmi sottesi al funzionamento della piattaforma. Tale blocco, della durata circa di un mese, si è risolto grazie a un rapido adeguamento da parte di OpenAI che ha aggiornato l'informativa e introdotto il diritto di opporsi a che i propri dati personali siano utilizzati per l'addestramento degli algoritmi[7].

Questo episodio si pone in linea con un'interpretazione particolarmente rigida del Regolamento Generale sulla Protezione dei Dati (GDPR 2016/679) dell'Unione Europea. Entrato in vigore il 25 maggio 2018, e introdotto in Italia attraverso il Decreto Legislativo 10 agosto 2018, n. 101, che ha adeguato la normativa nazionale alle disposizioni del GDPR, stabilisce le regole per il trattamento dei dati personali, inclusi i dati sanitari, e si applica a tutte le organizzazioni che operano nell'UE o che trattano i dati dei cittadini europei. I principi chiave del GDPR includono:

1. **liceità, correttezza e trasparenza: i dati personali devono essere trattati in modo lecito, corretto e trasparente nei confronti dell'interessato**
2. **limitazione della finalità: i dati personali devono essere raccolti per finalità determinate, esplicite e legittime, e successivamente trattati in modo compatibile con tali finalità**
3. **minimizzazione dei dati: i dati personali devono essere adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per le quali sono trattati**
4. **esattezza: i dati personali devono essere esatti e, se necessario, aggiornati**
5. **limitazione della conservazione: i dati personali devono essere conservati in una forma che consenta l'identificazione degli interessati per un arco di tempo non superiore al conseguimento delle finalità per le quali sono trattati**
6. **integrità e riservatezza: i dati personali devono essere trattati in maniera da garantire un'adeguata sicurezza, compresa la protezione contro trattamenti non autorizzati o illeciti e contro la perdita, la distruzione o il danno accidentali.**

---

[6] https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870847

[7] https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9881490

Quando viene affrontato il tema dei dati sanitari, il GDPR presenta delle accortezze particolari. I dati sanitari sono considerati una categoria speciale di dati personali che richiede una protezione aggiuntiva:

1. **il GDPR definisce i dati relativi alla salute come "dati personali relativi alla salute fisica o mentale di una persona fisica, compresa la prestazione di servizi di assistenza sanitaria, che rivelano informazioni relative al suo stato di salute" (Articolo 4)**
2. **quando il trattamento si basa sul consenso deve essere esplicito, libero, specifico e informato. Gli interessati hanno il diritto di revocare il consenso in qualsiasi momento (Articolo 7)**
3. **gli interessati hanno il diritto di accedere ai propri dati sanitari, ottenerne una copia, rettificarli se inesatti e, in alcune circostanze, ottenerne la cancellazione o la limitazione del trattamento (Articoli 15-18)**
4. **i titolari del trattamento devono attuare misure tecniche e organizzative appropriate per garantire un livello di sicurezza adeguato al rischio, come la pseudonimizzazione e la cifratura dei dati sanitari (Articolo 32)**
5. **il trattamento su larga scala di categorie particolari di dati, inclusi i dati sanitari, richiede una valutazione d'impatto sulla protezione dei dati prima del trattamento (Articolo 35)**
6. **i trasferimenti di dati sanitari al di fuori dell'UE/SEE sono consentiti solo se sono soddisfatte condizioni specifiche, come una decisione di adeguatezza o garanzie appropriate (Articoli 44-49).**

In sintesi, il GDPR impone obblighi stringenti per il trattamento dei dati sanitari, richiedendo una base giuridica valida, misure di sicurezza robuste e il rispetto dei diritti degli interessati. Tuttavia, il GDPR lascia ulteriore spazio agli Stati membri per ulteriori limitazioni a quanto delineato, consentendo un'interpretazione più restrittiva della norma.
La disciplina italiana ha storicamente affrontato il tema della protezione dei dati con un approccio conservativo, trasponendo la norma nel contesto italiano interpretandola in maniera più limitante di quanto delineato a livello europeo. Tuttavia, il recente Decreto PNRR, pubblicato in Gazzetta Ufficiale in data 2 maggio 2024, modifica il Codice della privacy (articolo 110-bis del Dlgs 196/2003) per allineare la normativa italiana al GDPR e per favorire la ricerca scientifica attraverso una norma che consente il riutilizzo dei dati personali a fini di ricerca scientifica anche senza il consenso degli interessati.
Questa notizia è stata accolta positivamente dalla comunità scientifica, che vede un primo segnale di apertura e di allineamento agli standard europei, rispetto a un'interpretazione che in passato ha avuto l'effetto di limitare la collaborazione e partecipazione a progetti europei e con partner internazionali.
Nel contesto europeo l'Artificial Intelligence Act (AI Act)[8] dell'Unione Europea, approvato nel marzo 2024 e pienamente operativo entro 48 mesi[9], regolamenta lo sviluppo, l'implementazione e l'uso dei sistemi di intelligenza artificiale all'interno dell'UE. L'AI Act non menziona specificamente gli LLM ma

[8] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

[9] https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law

classifica i sistemi di intelligenza artificiale utilizzati nell'assistenza sanitaria applicazioni "ad alto rischio". Ciò significa che tali sistemi saranno soggetti a requisiti rigorosi e a supervisione prima di poter essere immessi sul mercato o messi in servizio.
I fornitori di sistemi di intelligenza artificiale ad alto rischio devono garantire che i loro sistemi siano conformi ai requisiti stabiliti nell'AI Act, che includono:

- **stabilire e mantenere un sistema di gestione del rischio**
- **garantire che dataset di alta qualità siano utilizzati per l'addestramento, la validazione e il collaudo dei sistemi**
- **fornire una documentazione tecnica dettagliata e garantire una conservazione dei registri**
- **garantire la trasparenza e fornire informazioni chiare agli utenti**
- **implementare misure di supervisione umana sul funzionamento di questi sistemi**
- **garantire robustezza, accuratezza e sicurezza informatica.**

L'AI Act enfatizza anche l'importanza della governance dei dati, richiedendo che i set di dati di addestramento, convalida e collaudo siano pertinenti, rappresentativi, privi di errori e completi. Questo è particolarmente importante per i sistemi di intelligenza artificiale utilizzati nell'assistenza sanitaria, poiché dati distorti o imprecisi potrebbero portare a risultati dannosi per i pazienti. Inoltre, l'AI Act sottolinea la necessità di trasparenza e spiegabilità nei sistemi di intelligenza artificiale ad alto rischio. I fornitori devono garantire che il funzionamento dei loro sistemi di intelligenza artificiale sia sufficientemente trasparente da consentire agli utenti di interpretare l'output del sistema e di utilizzarlo in modo appropriato.
Come in passato, l'UE delinea i requisiti a un livello generale, permettendo un'interpretazione della norma ai singoli Stati membri.
Sarà necessario, per quanto complesso, definire standard per la valutazione del rischio e standard per la mitigazione del rischio che rispondano alle sfide uniche presentate dai sistemi di intelligenza artificiale quali gli LMM.

Il Disegno di legge italiano sull'intelligenza artificiale approvato il 23 aprile 2024 definisce alcuni principi fondamentali per l'uso dell'intelligenza artificiale nel settore sanitario e non solo. I principi fondamentali che il Disegno di legge intende garantire includono la protezione dei dati personali, l'equità, il diritto all'informazione, l'accessibilità, l'autonomia, la sicurezza e l'inclusione sociale. Il testo legislativo sottolinea che l'intelligenza artificiale non può sostituire il clinico; piuttosto, deve servire come supporto ai processi di prevenzione, diagnosi e scelta terapeutica, lasciando la decisione finale al giudizio del medico. Inoltre, è prevista la protezione dei cittadini attraverso l'aggiornamento e la verifica continua delle applicazioni di intelligenza artificiale.
Nell'Articolo 7, i principi chiave sono definiti come segue:

1. **"l'utilizzo di sistemi di intelligenza artificiale contribuisce al miglioramento del sistema sanitario e alla prevenzione e cura delle malattie, nel rispetto dei diritti, delle libertà e degli interessi della persona, anche in materia di protezione dei dati personali"**

2. **"l'introduzione di sistemi di intelligenza artificiale nel sistema sanitario non può selezionare e condizionare l'accesso alle prestazioni sanitarie con criteri discriminatori"**
3. **"l'interessato ha diritto di essere informato circa l'utilizzo di tecnologie di intelligenza artificiale e sui vantaggi, in termini diagnostici e terapeutici, derivanti dall'utilizzo delle nuove tecnologie, nonché di ricevere informazioni sulla logica decisionale utilizzata"**
4. **"la presente legge promuove lo sviluppo, lo studio e la diffusione di sistemi di intelligenza artificiale che migliorano le condizioni di vita delle persone con disabilità, agevolano l'accessibilità, l'autonomia, la sicurezza e i processi di inclusione sociale delle medesime persone anche ai fini dell'elaborazione del progetto di vita"**
5. **"i sistemi di intelligenza artificiale nell'ambito sanitario costituiscono un supporto nei processi di prevenzione, diagnosi, cura e scelta terapeutica, lasciando impregiudicata la decisione, che è sempre rimessa alla professione medica"**
6. **"i sistemi di intelligenza artificiale utilizzati nell'ambito sanitario e i relativi dati impiegati devono essere affidabili e periodicamente verificati e aggiornati al fine di minimizzare il rischio di errori".**

Per quanto riguarda la ricerca e la sperimentazione nell'ambito sanitario, delineata nell'Articolo 8, "i trattamenti di dati, anche personali, eseguiti da soggetti pubblici e privati senza scopo di lucro per la ricerca e la sperimentazione scientifica nella realizzazione di sistemi di intelligenza artificiale per finalità di prevenzione, diagnosi e cura, nonché di salute pubblica, sono dichiarati di rilevante interesse pubblico in attuazione dell'articolo 32 della Costituzione" e nel rispetto di quanto previsto nel GDPR. Per queste finalità "è sempre autorizzato l'uso secondario di dati personali privi degli elementi identificativi diretti", tenendo comunque in conto l'obbligo di informativa dell'interessato. Questo articolo rappresenta un passo importante per facilitare la ricerca e l'utilizzo dei dati esistenti nel settore sanitario, favorendo lo sviluppo di soluzioni innovative basate sull'IA che possono migliorare la qualità dell'assistenza e l'efficienza del sistema sanitario. Sarebbe auspicabile che in futuro anche le entità a scopo di lucro siano incluse in questa previsione, a condizione che garantiscano lo stesso livello di fiducia richiesto agli enti pubblici e privati senza scopo di lucro. Ciò potrebbe stimolare ulteriormente l'innovazione e la collaborazione tra il settore pubblico e quello privato, apportando competenze e risorse aggiuntive per lo sviluppo di soluzioni di intelligenza artificiale all'avanguardia in ambito sanitario. Naturalmente, sarà fondamentale che tutti i soggetti coinvolti adottino le necessarie precauzioni per proteggere la privacy e la sicurezza dei dati personali, garantendo la trasparenza e il rispetto dei diritti degli interessati.

L'Articolo 9 dettaglia le disposizioni relative all'uso di una piattaforma di intelligenza artificiale destinata a supportare le attività di cura, con un focus particolare sull'assistenza territoriale. Questa piattaforma, progettata, realizzata e gestita dall'Agenzia Nazionale per i Servizi Sanitari Regionali (Agenas) in qualità di Agenzia nazionale per la sanità digitale, è pensata per lavorare a supporto di professionisti sanitari, medici e utenti, fornendo "suggerimenti non vincolanti". È pertanto evidente che questa piattaforma si fonderà sull'utilizzo di LMM, con le opportunità e criticità paventate nei paragrafi precedenti di questo documento.
Tuttavia, lo stesso articolo enuncia che "dall'attuazione del presente articolo non devono derivare

nuovi o maggiori oneri per la finanza pubblica. L'Agenas provvede alle attività di cui al presente articolo con le risorse umane, strumentali e finanziarie disponibili a legislazione vigente". Qualora raggiunto, si tratterebbe di un traguardo significativo e certamente meritevole di lode.

Conclusioni

Gli LMM offrono notevoli opportunità per migliorare l'assistenza sanitaria, tuttavia è essenziale una gestione attenta dei rischi attraverso regolamentazioni adeguate, formazione continua e un impegno verso l'innovazione responsabile per proteggere la sicurezza dei pazienti e l'integrità dei sistemi sanitari. Il presente documento sottolinea l'importanza di un approccio collaborativo, globale e orientato all'etica nella governance dell'intelligenza artificiale, evidenziando l'urgenza di una regolamentazione efficace e responsabile per guidare lo sviluppo e l'uso delle tecnologie emergenti in modo sostenibile e inclusivo per tutta l'umanità. La SIIAM, riconoscendo l'importanza di adottare i più alti standard tecnici e qualitativi, nonché etici e di governance, contribuisce al progresso nel campo dell'implementazione dell'intelligenza artificiale in medicina, integrando quanto delineato dal documento OMS con il particolare contesto europeo e italiano. Pur riconoscendo come si tratti del primo passo di un lungo percorso, questo documento si pone come punto di riferimento per azioni future a livello nazionale e internazionale.

Abbreviazioni

IA	intelligenza artificiale
LLM	Large Language Model (modelli linguistici di grandi dimensioni)
LMM	Large Multi-Modal Model (modelli multimodali di grandi dimensioni)

Executive Summary

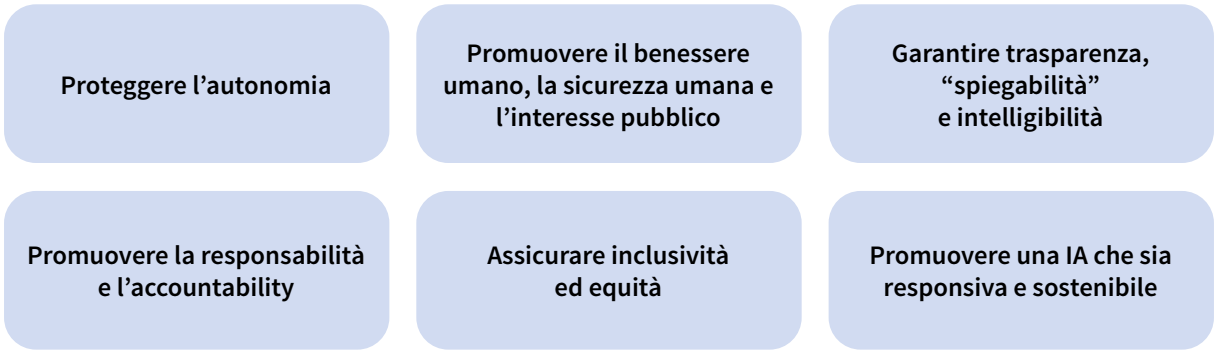
Il termine intelligenza artificiale (IA) fa riferimento alla capacità degli algoritmi, integrati all'interno di sistemi e strumenti, di apprendere dai dati in modo da eseguire compiti automatizzati senza l'esplicita programmazione di ogni passaggio da parte di un essere umano. L'IA generativa è una categoria di tecniche di IA i cui algoritmi vengono addestrati su dataset che possono essere utilizzati per generare nuovi contenuti, quali testi, immagini o video.
Queste linee guida (dell'Organizzazione Mondiale della Sanità, OMS, ndr) riguardano una particolare categoria di IA generativa: i modelli multimodali di grandi dimensioni (Large Multi-modal Model, LMM). Questi sono in grado di accettare una o più tipologie di input di dati e generare output diversi, non limitati al tipo di dati in ingresso all'algoritmo. Si prevede che gli LMM possano avere ampio utilizzo e numerose applicazioni in ambito sanitario, nella ricerca scientifica, nella sanità pubblica, e nello sviluppo di farmaci. Gli LMM sono anche conosciuti come "modelli di base per finalità generali" (general-purpose foundation models), sebbene non sia ancora stata dimostrata la loro efficacia in un'ampia gamma di compiti e obiettivi.

A oggi, gli LMM sono diventati il prodotto commerciale adottato più velocemente nella storia umana, e sono attraenti perché facilitano l'interazione tra uomo e computer, mimando la comunicazione umana e generando risposte, a domande o a input di dati, che appaiono reali, umane e credibili. A seguito della rapida adozione degli LMM da parte dei consumatori, e in vista del loro potenziale nel rivoluzionare molteplici servizi sociali e settori economici, diverse grandi aziende tecnologiche, start-up e governi stanno investendo e competendo tra loro per guidare lo sviluppo dell'IA generativa.

Nel 2021, l'OMS ha pubblicato delle linee guida complete [1] sull'etica e la governance dell'IA per la salute. L'OMS ha consultato 20 esperti di IA, che hanno identificato sia i potenziali benefici, sia i potenziali rischi connessi all'uso dell'IA nell'assistenza sanitaria, e hanno elaborato, con il metodo del consenso, sei principi guida generali, da considerare nelle politiche e nelle pratiche di governi, sviluppatori e fornitori che utilizzano l'IA. Tali principi dovrebbero guidare lo sviluppo e l'implementazione dell'IA nell'assistenza sanitaria da parte di una vasta gamma di stakeholder, inclusi governi nazionali, agenzie del settore pubblico, ricercatori, aziende e utilizzatori. I principi sono i seguenti:

1. **proteggere l'autonomia**
2. **promuovere il benessere umano, la sicurezza umana e l'interesse pubblico**
3. **garantire trasparenza, "spiegabilità" e intelligibilità**
4. **promuovere la responsabilità e l'accountability**
5. **assicurare inclusività ed equità**
6. **promuovere un'IA che sia responsiva e sostenibile (Figura 1, pagina 15).**

Figura 1. *Principi etici per l'uso dell'IA in ambito sanitario*



L'OMS pubblica queste linee guida per assistere gli stati membri nel mappare i benefici e le sfide associate all'utilizzo degli LMM per la salute, e nello sviluppare politiche e pratiche per un loro sviluppo e utilizzo appropriati. Le linee guida includono raccomandazioni per la governance, all'interno delle aziende, da parte dei governi e attraverso la collaborazione internazionale, in linea con i principi guida. I principi guida e le raccomandazioni, che tengono conto degli unici mezzi attraverso i quali gli esseri umani possono utilizzare l'IA generativa per la salute, costituiscono le basi di queste linee guida.

Applicazioni, sfide e rischi dei modelli multimodali di grandi dimensioni

Le potenziali applicazioni degli LMM in ambito sanitario sono simili a quelle di altre forme di IA, tuttavia il loro utilizzo e la loro accessibilità rappresentano una novità, con annessi nuovi benefici e rischi che la società, i sistemi sanitari e gli utenti finali potrebbero non essere ancora completamente pronti ad affrontare. La **Tabella 1** (pagina 16) riassume le principali applicazioni degli LMM e i loro potenziali benefici e rischi.
I rischi sistemici associati all'uso degli LMM includono quelli che potrebbero influenzare i sistemi sanitari (**Tabella 2**, pagina 17).

Con l'utilizzo degli LMM potrebbero emergere rischi regolatori e sistemici più ampi. Una preoccupazione (esaminata da diverse autorità di protezione dei dati) riguarda il rispetto degli attuali vincoli legali o regolatori nel loro utilizzo, compresi gli obblighi internazionali sui diritti umani e le normative nazionali sulla protezione dei dati. Gli algoritmi potrebbero non rispettare tali vincoli a causa del modo in cui vengono raccolti i dati usati per addestrare gli LMM, della gestione ed elaborazione dei dati raccolti (o inseriti negli LMM dagli utenti finali), della trasparenza e della responsabilità delle entità che sviluppano gli LMM e della possibilità che gli LMM abbiano "allucinazioni". Gli LMM potrebbero anche non essere conformi alle leggi sulla protezione dei consumatori.

Rischi per la società più ampi, associati al crescente utilizzo degli LMM (includendo, e andando oltre l'uso di tali algoritmi nell'assistenza sanitaria), comprendono il fatto che gli LMM siano spesso sviluppati e distribuiti da grandi aziende tecnologiche e sono dovuti, in parte, alla significativa mole di

Tabella 1. *Potenziali benefici e rischi in diversi utilizzi degli LMM in ambito sanitario*

Uso	Benefici potenziali o attesi	Rischi potenziali
Diagnosi e cura	• Assistere nella gestione di casi complessi e nella revisione delle diagnosi • Ridurre il carico di lavoro comunicativo dei professionisti sanitari (liberazione dalla tastiera) • Fornire nuovi spunti e report da varie forme non strutturate di dati sanitari	• Risposte imprecise, incomplete o false • Scarsa qualità dei dati di addestramento dell'LMM • Bias (dei dati di addestramento e delle risposte) • Bias di automazione • Perdita delle competenze dei professionisti sanitari • Consenso informato (dei pazienti)
Utilizzo da parte del paziente	• Generare informazioni per migliorare la comprensione di una condizione medica (come paziente o come caregiver) • Assistente virtuale per la salute • Arruolamento in sperimentazioni cliniche	• Dichiarazioni imprecise, incomplete o false • Manipolazione • Privacy • Minore interazione tra professionisti sanitari e pazienti • Ingiustizia epistemica • Rischio di erogazione di cure al di fuori del sistema sanitario
Compiti d'ufficio e amministrativi	• Assistere nella gestione della documentazione e della modulistica necessaria per l'assistenza sanitaria • Assistere nella traduzione linguistica • Compilare le cartelle cliniche elettroniche • Redigere note cliniche dopo una visita del paziente	• Imprecisioni ed errori • Risposte non coerenti rispetto alle richieste (prompt)
Educazione medica e infermieristica	• Testi dinamici adattati alle esigenze di ciascuno studente • Conversazione simulata per migliorare la comunicazione e fare esperienze di pratica in diversi scenari e con diversi pazienti • Risposte alle domande accompagnate da ragionamenti sequenziali	• Contribuire al bias di automazione • Errori o informazioni false compromettono la qualità dell'educazione medica • Carico didattico aggiuntivo per l'apprendimento delle competenze digitali
Ricerca scientifica e sviluppo di farmaci	• Generare interpretazioni da dati scientifici e ricerche • Generare i testi per articoli scientifici, sottomissione di manoscritti o peer-review • Analizzare e riassumere i dati per una ricerca • Revisione di bozze • Ideazione di farmaci innovativi	• Non è possibile ritenere gli algoritmi responsabili del contenuto • Gli algoritmi hanno un pregiudizio orientato verso le prospettive dei Paesi ad alto reddito • Generare informazioni e/o riferimenti che non esistono • Sottovalutare i principi fondamentali della ricerca scientifica, come la revisione tra pari • Esacerbare le disuguaglianze di accesso alla conoscenza scientifica

Tabella 2. *Rischi associati all'uso degli LMM per i sistemi sanitari*

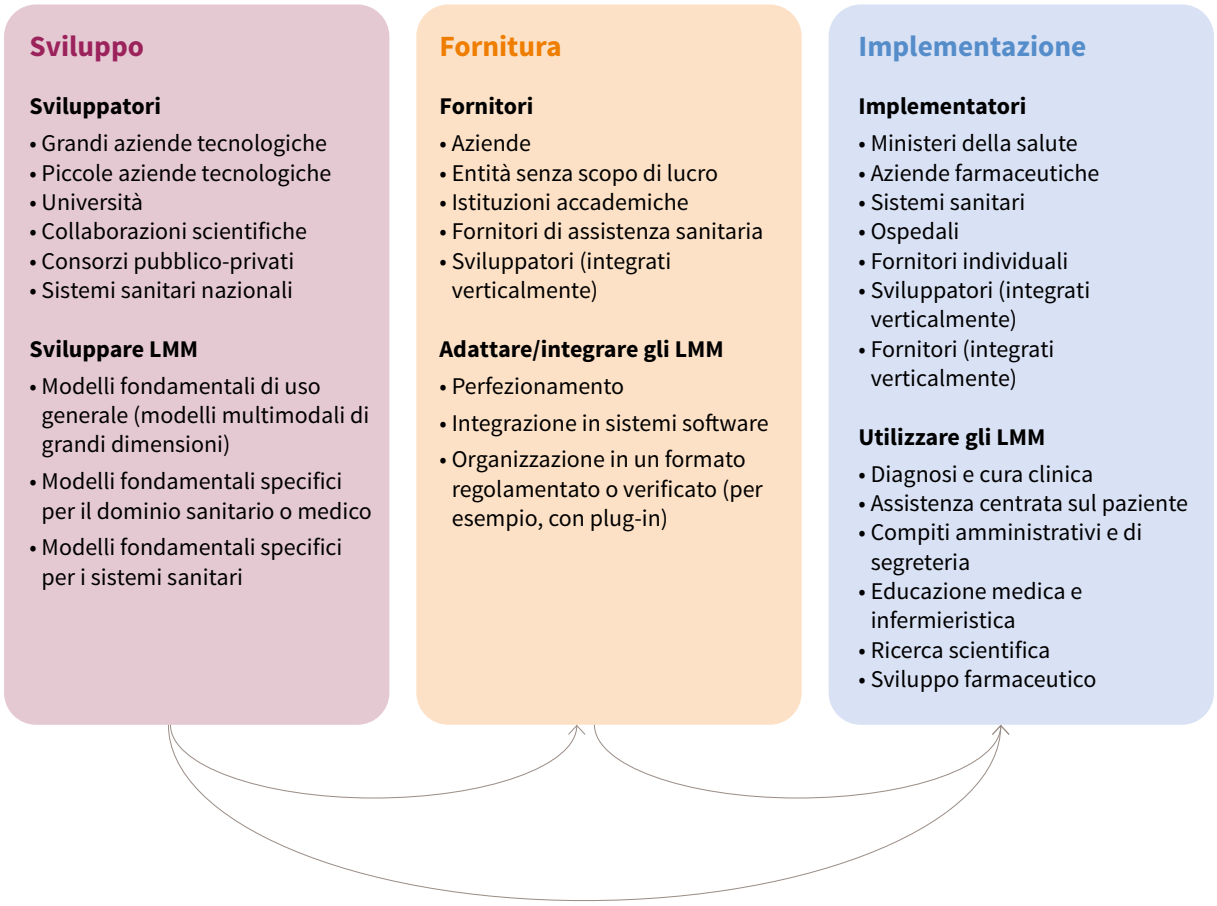
Tipo di rischio	Descrizione
Sovrastima dei benefici degli LMM	Potrebbe esserci una tendenza al ricorso alla tecnologia per avere risposte, o una sovrastima dei benefici degli LMM, trascurando o minimizzando le sfide nel loro utilizzo, compresa la sicurezza, l'efficacia e l'utilità
Accessibilità e convenienza	L'equità di accesso agli LMM può essere carente per diverse ragioni, tra cui il divario digitale e i costi di abbonamento per accedere agli LMM
Bias di sistema	L'uso di dataset sempre più grandi potrebbe aumentare i bias presenti negli LMM, che potrebbero diventare pervasivi in tutto un sistema sanitario
Impatti sull'occupazione	L'uso degli LMM potrebbe portare a perdite di posti di lavoro in alcuni Paesi e richiedere ai professionisti sanitari di riqualificarsi e adattarsi all'uso degli LMM. La codifica e il filtraggio dei dati possono portare a salari bassi e a stress psicologico
Dipendenza dei sistemi sanitari da LMM inadatti o centrati sulla malattia	La dipendenza dagli LMM potrebbe rendere i sistemi sanitari vulnerabili qualora gli LMM non dovessero venire mantenuti o (per quel che riguarda i Paesi a basso e medio reddito) venissero aggiornati solo per un uso nei Paesi ad alto reddito. Inoltre, la mancanza di conservazione e protezione della privacy e della riservatezza potrebbe minare la fiducia nei sistemi sanitari da parte delle persone che non sono sicure che la loro privacy sarà rispettata
Rischi per la sicurezza informatica	Attacchi dannosi o hackeraggi potrebbero compromettere la sicurezza e la fiducia nell'uso degli LMM in ambito sanitario

risorse informatiche, umane e finanziarie richieste per lo sviluppo degli LMM. Ciò può rafforzare la posizione di queste grandi aziende rispetto alle piccole imprese e ai governi nazionali per quanto riguarda lo sviluppo e l'uso dell'IA, incluso il focus della ricerca sull'IA nei settori pubblico e privato. Preoccupazioni aggiuntive, legate al potenziale dominio di mercato delle grandi aziende tecnologiche, riguardano un possibile insufficiente loro impegno all'etica e alla trasparenza. Nuovi accordi volontari tra le aziende e gli enti governativi potrebbero mitigare alcuni rischi nel breve termine, ma non devono essere considerati un'alternativa alla supervisione governativa che potrebbe essere messa in atto.

Un altro rischio per la società è legato all'impronta di carbonio e idrica degli LMM, i quali, come altre forme di IA, richiedono significative risorse energetiche e contribuiscono all'aumento dell'impronta idrica. Mentre gli LMM e altre forme di IA possono fornire importanti benefici sociali, l'aumento dell'impronta di carbonio potrebbe diventare un grande determinante del cambiamento climatico, e l'aumento del consumo di acqua può avere un ulteriore impatto negativo nelle comunità già afflitte da carenza idrica.

Un altro rischio per la società associato all'utilizzo degli LMM è che, fornendo risposte plausibili, che vengono considerate sempre più come una fonte di conoscenza, essi potrebbero minare l'autorità epistemica umana, inclusi i domini dell'assistenza sanitaria, della scienza e della medicina.

Figura 2. *Catena del valore dello sviluppo, della fornitura e dell'implementazione degli LMM*



Etica e governance degli LMM nell'assistenza sanitaria e nella medicina

Gli LMM possono essere considerati prodotti di una serie (o catena) di decisioni riguardanti la programmazione e lo sviluppo del prodotto da parte di uno o più attori (**Figura 2**). Le decisioni prese in ogni fase della catena del valore dell'IA possono avere conseguenze sia dirette sia indirette su coloro che partecipano a valle allo sviluppo, alla disseminazione e all'uso degli LMM. Le decisioni possono essere influenzate e regolamentate dai governi attraverso l'emanazione e l'applicazione di leggi e politiche a livello nazionale, regionale e globale.

La catena del valore dell'IA spesso inizia in una grande azienda tecnologica, definita "sviluppatore" in queste linee guida. Lo sviluppatore potrebbe anche essere un'università, una piccola azienda tecnologica, un sistema sanitario nazionale, un consorzio pubblico-privato o un'altra entità che disponga delle risorse e della capacità di utilizzare diversi input, che costituiscono "l'infrastruttura dell'IA", come dati, potenza di calcolo e competenza di IA, per sviluppare "modelli di base per finalità generali" (un termine utilizzato per descrivere gli LMM in ambito legislativo e regolatorio). Questi modelli di uso generale possono essere utilizzati direttamente per svolgere vari compiti, spesso non previsti o predeterminati, inclusi quelli legati all'assistenza sanitaria. Diversi modelli sono invece addestrati specificamente per l'uso nell'assistenza sanitaria e nella medicina.

Un modello di base per finalità generali può essere utilizzato da una terza parte (un "fornitore") attraverso un'interfaccia di programmazione attiva per uno scopo o un uso specifico.
Questo comporta:

1. **il perfezionamento (fine-tuning) di un nuovo LMM, che potrebbe richiedere un ulteriore addestramento del modello fondativo**
2. **l'integrazione dell'LMM in applicazioni o in un sistema software più ampio per fornire un servizio agli utenti**
3. **l'integrazione di componenti noti come "plug-in" per incanalare, filtrare e organizzare gli LMM in versioni formali o regolamentate per generare risultati "digeribili".**[1]

Successivamente, il fornitore può commercializzare un prodotto o un servizio basato sull'LMM a un cliente (o "implementatore"), come un ministero della salute, un sistema sanitario, un ospedale, un'azienda farmaceutica o anche una singola persona, come un professionista che fornisce assistenza sanitaria. Il cliente che acquista o prende in licenza il prodotto o l'applicazione può poi utilizzarlo direttamente per i pazienti, i professionisti sanitari, altre entità del sistema sanitario, persone comuni o nella propria attività. La catena del valore può quindi essere "integrata verticalmente", in modo che un'azienda (o un'altra entità, come un sistema sanitario nazionale), che raccoglie dati e addestra un modello di base, possa modificare l'LMM per un uso specifico e fornire l'applicazione direttamente ai propri utenti.

La governance è un mezzo per preservare e salvaguardare i principi etici e gli obblighi rispetto ai diritti umani attraverso leggi e politiche preesistenti e attraverso nuove leggi, norme, codici interni di comportamento e procedure per gli sviluppatori, ma anche accordi e quadri normativi internazionali. Un modo per inquadrare e definire la governance degli LMM è seguire le tre fasi della catena del valore dell'IA:

1. **la progettazione e lo sviluppo di modelli di base di uso generale o LMM**
2. **la fornitura di un servizio, applicazione o prodotto basato su un LMM**
3. **l'implementazione di un servizio o di un'applicazione sanitaria.**

In questa linea guida, ogni fase è esaminata rispetto a tre aree di approfondimento:

- **quali rischi (descritti sopra) devono essere considerati in ogni fase della catena del valore, e quali attori sono meglio posizionati per affrontare questi rischi?**
- **che cosa può fare un attore rilevante per affrontare i rischi, e quali principi etici devono essere rispettati?**
- **qual è il ruolo dei governi, incluse leggi, politiche e regolamenti?**

Alcuni rischi possono essere affrontati in ogni fase della catena del valore dell'IA, e alcuni attori sono probabilmente più importanti di altri nel mitigare determinati rischi e nel promuovere gli aspetti

[1] Communication from Leong Tze-Yun, WHO expert on the ethics and governance of AI for health.

etici. Sebbene sia probabile che si abbiano disaccordi e tensione riguardo ai profili di responsabilità tra sviluppatori, fornitori e implementatori, ci sono aree ben definite in cui ogni attore è chiaramente posizionato o è l'unica entità capace di affrontare un rischio potenziale o effettivo.

Progettazione e sviluppo di modelli multimodali di grandi dimensioni (LMM)

Durante la progettazione e lo sviluppo di modelli di base a uso generale, la responsabilità ricade sugli sviluppatori. I governi hanno il compito di stabilire leggi e standard per favorire o proibire certe pratiche. La **Sezione 4** di queste linee guida fornisce raccomandazioni per affrontare i rischi e massimizzare i benefici durante lo sviluppo degli LMM.

Fornitura dei modelli di base per finalità generali (LMM)

Durante la fornitura di un servizio o di un'applicazione, i governi sono responsabili di definire i requisiti e gli obblighi sia degli sviluppatori sia dei fornitori per affrontare rischi specifici associati ai sistemi basati sull'IA usati in ambito sanitario. La **Sezione 5** di queste linee guida fornisce raccomandazioni per affrontare i rischi e massimizzare i benefici relativi alla fornitura di servizi e applicazioni in ambito sanitario basati sugli LMM.

Implementazione di modelli multimodali di grandi dimensioni (LMM)

Anche se nell'ambito dello sviluppo e della fornitura di un LMM vengono applicate le leggi, le politiche e le pratiche etiche rilevanti, alcuni rischi si possono verificare durante il loro utilizzo, in parte a causa dell'imprevedibilità degli LMM e delle risposte che forniscono, della possibilità che un utente applichi un LMM in un modo che né lo sviluppatore né il fornitore avevano previsto, e in parte perché gli output degli LMM possono cambiare nel tempo. La **Sezione 6** di queste linee guida fornisce raccomandazioni rispetto ai rischi e alle sfide che dovrebbero essere affrontate durante l'uso degli LMM e delle relative applicazioni.

Responsabilità legate agli LMM

Con il crescente uso degli LMM nell'assistenza sanitaria e in medicina saranno inevitabili errori, usi impropri con conseguenze negative per le persone. Pertanto, le norme sulla responsabilità potrebbero garantire che gli utenti danneggiati da un LMM siano adeguatamente risarciti o ottengano altre forme di compensazione, riducendo l'onere della prova da parte di una persona danneggiata, assicurando che sia adeguatamente ed equamente risarcita.

I governi possono fare ciò introducendo una presunzione di causalità. Potrebbero anche considerare l'introduzione di un rigoroso standard di responsabilità per qualsiasi danno che derivi dall'implementazione di un LMM. Tuttavia, sebbene norme stringenti sulla responsabilità possano garantire il risarcimento per coloro che subiscono danni, potrebbero anche scoraggiare l'uso di LMM sempre più sofisticati. I governi potrebbero anche considerare fondi di compensazione "senza colpa" e "senza responsabilità".

Governance internazionale degli LMM

I governi devono collaborare per costruire nuove strutture istituzionali e regole per garantire che la governance internazionale mantenga il passo con la globalizzazione di queste tecnologie. Dovrebbero anche garantire una cooperazione e collaborazione più forte all'interno dell'ONU per rispondere alle opportunità e alle sfide legate all'implementazione dell'IA in ambito sanitario, così come alle sue più ampie applicazioni nella società e nell'economia.

È necessaria una governance internazionale per garantire che tutti i governi nazionali siano responsabili dei loro investimenti e della loro partecipazione nello sviluppo e nell'implementazione

di sistemi basati sull'IA, e che gli stessi governi introducano regolamentazioni appropriate che rispettino i principi etici, i diritti umani e il diritto internazionale. Una governance internazionale potrà anche garantire che le aziende sviluppino e implementino LMM capaci di soddisfare adeguati standard di sicurezza ed efficacia e rispettino i principi etici e i diritti umani. I governi nazionali dovrebbero anche evitare di introdurre regolamentazioni che offrano un vantaggio competitivo o uno svantaggio, sia per le aziende, sia per sé.

Affinché la governance internazionale risulti significativa, tali regole devono essere plasmate e decise da tutti i Paesi coinvolti, e non solo dai Paesi ad alto reddito (e dalle aziende tecnologiche che lavorano con i governi dei Paesi ad alto reddito). Una governance internazionale può richiedere che tutti i portatori di interesse cooperino attraverso una rete multilaterale, come proposto nel 2019 dal segretario generale dell'ONU, che riunirebbe le Nazioni unite, le istituzioni finanziarie internazionali, le organizzazioni regionali, i blocchi commerciali e altri attori, inclusa la società civile, i comuni, le aziende, le autorità locali e i giovani, per lavorare in modo più collaborativo, efficace e inclusivo.

Etica e governance dell'intelligenza artificiale per la salute: modelli multimodali di grandi dimensioni

Rischi da affrontare

Che cosa può essere fatto e da chi

Fase di sviluppo

- Bias
- Privacy
- Lavoro e occupazione
- Impronta idrica e di CO_2
- Informazioni false e disinformazione
- Sicurezza e cyber-sicurezza
- Autorità epistemologica umana
- Controllo esclusivo degli LLM

Azioni degli sviluppatori

- Certificazione/formazione per programmatori
- Valutazioni di impatto sulla protezione dei dati
- Dati di training raccolti con regole di protezione dati "best-practice"
- Dati di training aggiornati, attuali e appropriati al contesto
- Garantire la trasparenza dei dati di training
- Salari equi e sostegno ai lavoratori
- Coinvolgere stakeholder diversificati nella progettazione
- Progettare per migliorare l'accuratezza e la prevedibilità
- Progettare per migliorare l'efficienza energetica dei modelli

Azioni governative

- Avere e applicare leggi forti sulla protezione dei dati
- Emissione di profili di prodotto target
- Progettare per valori basati su principi di consenso e norme etiche
- Imporre risultati (prevedibilità, interpretabilità, correggibilità, sicurezza, sicurezza informatica)
- Promuovere gli LMM open source con finanziamenti
- Condurre audit durante lo sviluppo iniziale dell'IA

Fase di fornitura

- Pregiudizio a livello di sistema
- Manipolazione
- Bias di automazione
- Informazioni false e disinformazione
- Privacy

Azioni governative

- Istituire un'agenzia regolatoria per valutare e approvare gli LMM per la salute
- Richiedere trasparenza, inclusi codice sorgente e dati di input
- Applicare le leggi sulla protezione dei dati per i dati inseriti dagli utenti
- Imporre standard etici e di diritti umani, indipendentemente dal rischio o dal beneficio
- Promulgare leggi che richiedono valutazioni di impatto condotte da terze parti e divulgate pubblicamente
- Richiedere prove di performance e conformità con le normative sui dispositivi medici
- Proibire l'uso sperimentale al di fuori di ambiti di ricerca; esplorare sandbox normativi per test controllati
- Applicare leggi sulla protezione dei consumatori per prevenire impatti negativi sugli utenti finali e sui pazienti

Fase di implementazione

- Inaccuratezza o false risposte
- Bias
- Privacy
- Accessibilità
- Lavoro e occupazione
- Bias di automazione
- Qualità dell'interazione paziente-professionista
- Degrado delle competenze

Azioni degli implementatori

- Evitare l'uso degli LMM in contesti inappropriati
- Comunicare rischi, errori e danni noti con avvertenze chiare
- Garantire accessibilità e disponibilità assicurando che prezzi e lingue offerte siano inclusivi

Azioni governative

- Imporre audit e valutazioni di impatto post rilascio indipendenti per l'implementazione degli LMM
- Rendere responsabili gli sviluppatori o i fornitori per informazioni false o tossiche
- Applicare divulgazioni operative, incluse documentazioni tecniche
- Formare i lavoratori sanitari sul processo decisionale degli LMM, evitando pregiudizi, coinvolgimento dei pazienti e rischi per la sicurezza informatica
- Facilitare la partecipazione pubblica attraverso collegi di supervisione umana per garantire un uso appropriato
- Coinvolgere il pubblico per comprendere la condivisione dei dati, valutare l'accettabilità sociale e culturale, migliorare l'alfabetizzazione sull'IA e valutare gli usi accettabili degli LMM
- Utilizzare l'autorità di approvvigionamento per incoraggiare la trasparenza e pratiche responsabili da parte degli attori della catena del valore

1. Introduzione

Questa guida si occupa delle nuove applicazioni dei modelli multimodali di grandi dimensioni (LMM) in ambito sanitario[1], compresi i potenziali benefici e i rischi dell'utilizzo degli LMM nell'assistenza sanitaria e in medicina, nonché degli approcci alla governance degli LMM per garantire il rispetto delle linee guida e degli obblighi in materia di etica, diritti umani e sicurezza. Questa guida riprende la linea guida dell'OMS emessa nel giugno 2021 "Ethics and governance of artificial intelligence for health" [1], la quale ha affrontato le sfide etiche e i rischi dell'uso dell'intelligenza artificiale (IA) in campo sanitario, ha identificato sei principi per garantire che l'IA sia utilizzata per il bene pubblico di tutti i Paesi e ha emesso raccomandazioni per migliorare la governance dell'IA per la salute al fine di massimizzare le potenzialità di questa tecnologia.

L'IA si riferisce alla capacità di algoritmi integrati in sistemi e strumenti di imparare dai dati per svolgere compiti automatizzati, senza la programmazione esplicita di ogni passaggio da parte di un essere umano. L'IA generativa è una categoria di tecniche di IA in cui i modelli di apprendimento automatico (machine learning) vengono utilizzati per addestrare tali algoritmi su insiemi di dati (dataset) al fine di creare nuovi output, quali testo, immagini, video e musica. I modelli di IA generativa apprendono schemi e strutture dai dati di addestramento e producono nuovi dati basati su previsioni fatte dai pattern appresi. I modelli di IA generativa possono essere migliorati attraverso l'apprendimento per rinforzo grazie al feedback umano, nel cui ambito gli operatori classificano le risposte fornite dai modelli dell'IA generativa così da addestrare gli algoritmi a processare output che massimizzino il punteggio che gli esseri umani vanno ad assegnare. L'IA generativa ha applicazioni potenziali in svariati campi che comprendono il design, la generazione di contenuti, la simulazione e la ricerca scientifica.

Grande attenzione è stata rivolta a un particolare tipo di IA generativa, i modelli linguistici di grandi dimensioni (Large Language Model, LLM), i quali ricevono un unico tipo di input in forma testuale e forniscono una risposta anch'essa in forma testuale. I modelli linguistici di grandi dimensioni sono un esempio dei modelli unimodali di grandi dimensioni, che costituiscono la base per il funzionamento delle primissime versioni dei chatbot che integrano questi modelli. Sebbene l'interazione con gli LLM simuli un dialogo, i modelli stessi non hanno alcuna concezione di ciò che stanno producendo. Questi, infatti, prevedono meramente la parola successiva in base alle parole precedenti, in base ai pattern appresi o in base alle combinazioni di parole [2].
Questo documento affronta l'uso crescente degli LMM (inclusi gli LLM), i quali, per essere impiegati nell'assistenza sanitaria e nella medicina, vengono addestrati con set di dati altamente eterogenei,

[1] Ai fini di questa guida, i termini "modelli multimodali di grandi dimensioni" e "modelli di base a uso generale" sono utilizzati in modo intercambiabile; quest'ultimo termine viene utilizzato soprattutto in ambito di governance. Tuttavia, non è ancora noto se gli LMM possano svolgere una vasta gamma di compiti a fini generali.

che vanno oltre il testo e includono dati provenienti da biosensori, dati genomici, epigenomici, proteomici, dati di imaging, clinici, sociali e ambientali [3]. Pertanto, gli LMM possono processare più di un tipo di input e generare output non limitati al tipo di dati inseriti. Gli LMM sono pensati per diverse applicazioni nell'ambito dell'assistenza sanitaria e dello sviluppo di farmaci.

Gli LMM si distinguono dalle forme precedenti di IA e machine learning. Mentre l'IA è già stata ampiamente integrata in molte applicazioni per consumatori, gli output della maggior parte degli algoritmi non richiedono né invitano la partecipazione dell'utente, tranne per forme rudimentali di IA integrate nelle piattaforme di social media che curano i contenuti generati dagli utenti per catturare l'attenzione [4]. Un'altra differenza tra gli LMM e altre forme di IA è la loro versatilità. I modelli di IA precedenti ed esistenti, compresi quelli per usi medici, sono progettati per compiti specifici e di conseguenza sono rigidi, possono eseguire solo compiti definiti nel set di addestramento [5] e non possono adattarsi o svolgere altre funzioni senza essere riaddestrati con un dataset diverso. Pertanto, anche se la Food and Drug Administration negli Stati Uniti ha approvato più di 500 modelli di IA per l'utilizzo nella pratica medica [5], la maggior parte è approvata solo per uno o due compiti molto specifici. Al contrario, gli LMM sono addestrati su vari dataset e possono essere utilizzati per numerosi compiti, compresi alcuni per i quali non sono stati esplicitamente addestrati [5].

Gli LMM di solito hanno un'interfaccia e un formato che facilitano le interazioni tra l'uomo e l'algoritmo e che può imitare la comunicazione umana, potendo quindi indurre gli utenti ad attribuire all'algoritmo qualità simili a quelle umane. Per questo motivo, il modo in cui vengono utilizzati gli LMM e i contenuti che generano e forniscono come risposte – che possono apparire generate da un essere umano – sono diversi da quelli di altre forme di IA e hanno contribuito a un utilizzo degli LMM senza precedenti. Inoltre, poiché le risposte che forniscono sembrano essere autorevoli, molti utenti le accettano acriticamente come corrette, anche se un LMM non può garantire una risposta corretta e non può integrare norme etiche o ragionamenti morali nelle risposte che genera. Mentre questa guida illustra i diversi modi in cui gli LMM vengono utilizzati (o potrebbero essere usati) nell'assistenza sanitaria e nella medicina, essi sono già impiegati in numerosi settori quali l'istruzione, la finanza, le comunicazioni e l'informatica.

Gli LMM possono essere considerati il risultato finale di una serie (o catena) di decisioni in materia di programmazione e sviluppo da parte di uno o più attori. Le decisioni prese in ogni fase della catena del valore dell'IA possono avere conseguenze dirette e indirette su coloro che partecipano allo sviluppo, alla distribuzione e all'uso degli LMM a valle. Le decisioni possono essere influenzate e regolate dai governi che promulgano e applicano leggi e politiche a livello nazionale, regionale e globale. La catena del valore dell'IA spesso inizia in una grande azienda tecnologica, definita "sviluppatore" in queste linee guida. Lo sviluppatore potrebbe anche essere un'università, una piccola azienda tecnologica, un sistema sanitario nazionale, un consorzio pubblico-privato o un'altra entità che disponga delle risorse e della capacità di utilizzare diversi input, che costituiscono "l'infrastruttura dell'IA", come dati, potenza di calcolo e competenza di IA, per sviluppare "modelli di base per finalità generali" (un termine utilizzato per descrivere gli LMM in ambito legislativo e regolatorio). Questi modelli di uso generale possono essere utilizzati direttamente per svolgere vari compiti, spesso non previsti

o predeterminati, inclusi quelli legati all'assistenza sanitaria. Diversi modelli sono invece addestrati specificamente per l'uso nell'assistenza sanitaria e nella medicina.
Un modello di base per finalità generali può essere utilizzato da una terza parte (un "fornitore") attraverso un'interfaccia di programmazione attiva per uno scopo o un uso specifico. Questo comporta:

1. **il perfezionamento (fine-tuning) di un nuovo LMM, che potrebbe richiedere un ulteriore addestramento del modello fondativo**
2. **l'integrazione dell'LMM in applicazioni o in un sistema software più ampio per fornire un servizio agli utenti**
3. **l'integrazione di componenti noti come "plug-in" per incanalare, filtrare e organizzare gli LMM in versioni formali o regolamentate per generare risultati "digeribili"[2].**

In seguito, il fornitore può commercializzare un prodotto o un servizio basato sull'LMM a un cliente (o a un "distributore"), come un ministero della salute, un servizio sanitario, un ospedale, una azienda farmaceutica o anche al singolo individuo, come un professionista sanitario. Il cliente che acquista o acquisisce in licenza il prodotto o l'applicazione può quindi utilizzarlo direttamente per i pazienti, per i professionisti sanitari, per altri enti del servizio sanitario, per persone comuni o nella propria azienda. La catena del valore può essere "integrata verticalmente ", in modo che un'azienda (o altra entità, come un servizio sanitario) che raccoglie dati e addestra un modello di base a uso generale possa modificare l'LMM per un uso particolare e fornire l'applicazione direttamente agli utenti.

L'OMS riconosce i grandi benefici che l'IA potrebbe apportare ai sistemi sanitari, tra cui il miglioramento della salute pubblica e il raggiungimento della copertura sanitaria universale. Tuttavia, come descritto nella linea guida OMS "Ethics and governance of artificial intelligence for health" [1], comporta significativi rischi che potrebbero minare la salute pubblica e mettere a repentaglio la dignità del singolo, la privacy e i diritti umani. Sebbene gli LMM siano relativamente nuovi, la velocità della loro adozione e diffusione ha portato l'OMS a fornire queste linee guida per garantire che vengano possibilmente utilizzati al meglio e in maniera sostenibile in tutto il mondo. L'OMS riconosce che queste linee guida vengono emesse in un momento in cui ci sono molte opinioni contrastanti sui potenziali benefici e rischi dell'IA, sui principi etici che dovrebbero essere applicati alla sua costruzione e al suo utilizzo e sugli approcci alla governance e alla regolamentazione. Poiché questa linea guida è stata pubblicata poco dopo le prime applicazioni degli LMM nell'assistenza sanitaria e prima che vengano rilasciati modelli più potenti, l'OMS aggiornerà le linee guida per tenere il passo con l'evoluzione rapida della tecnologia, come la società gestisce il suo utilizzo e le conseguenze sulla salute dell'uso degli LMM al di là dell'assistenza sanitaria e della medicina.

1.1 Rilevanza dei modelli multimodali di grandi dimensioni (LMM)

Gli LMM, sebbene relativamente recenti e non testati, hanno avuto un impatto enorme sulla società in vari ambiti, compresa l'assistenza sanitaria e la medicina. È stato stimato che ChatGPT, un modello linguistico di grandi dimensioni le cui versioni successive sono state rilasciate da un'azienda tec-

[2] Comunicazione del Dott. Leong Tze-Yun, esperto dell'OMS sull'etica e la governance dell'IA in sanità.

nologica statunitense, abbia raggiunto 100 milioni di utenti attivi mensili nel gennaio 2023, appena 2 mesi dopo il suo lancio. Tale risultato ha reso ChatGPT l'applicazione a licenza di largo consumo (consumer application) con più rapida crescita nella storia [6].
Numerose aziende stanno a oggi sviluppando LMM, o integrando gli LMM in applicazioni destinate ai consumatori, quali i motori di ricerca in Internet. Grandi aziende tecnologiche stanno rapidamente integrando gli LMM in quasi tutte le applicazioni, o ne stanno creando di nuove [7,8]. Nuove aziende, con milioni di dollari di investimenti privati, stanno sviluppando LMM [9]. Anche gli LMM open source stanno emergendo più rapidamente e a costi inferiori rispetto a quelli sviluppati dalle più grandi aziende, grazie alla disponibilità di tali piattaforme [10].

Anche se l'avvento degli LMM sta alimentando nuovi investimenti nel settore tecnologico e nuovi prodotti vengono rilasciati costantemente, alcune tra le aziende stesse ammettono di non comprendere appieno perché gli LMM generino determinate risposte [11]. Nonostante il ricorso a un addestramento tramite rinforzo con feedback umano (reinforcement learning), gli LMM possono generare output che non sono sempre prevedibili o controllati, compresi LMM che intraprendono "conversazioni" che possono essere sgradite agli utenti [12], o che pubblicano contenuti errati o comunque imprecisi ma altamente convincenti [13]. Tuttavia, gran parte del sostegno nei confronti degli LMM non risiede solo nell'entusiasmo per le loro funzionalità, ma anche in affermazioni non qualificate e poco oggettive facenti riferimento a prestazioni degli LMM in pubblicazioni non sottoposte a una revisione scientifica tra pari [14].

Gli LMM sono stati adottati rapidamente, anche se i dataset utilizzati per addestrarli non sono stati divulgati [15], rendendo così difficile o impossibile sapere se i dati sono affetti da bias, se sono stati acquisiti legalmente e in conformità con le normative sulla protezione dei dati e infine se i risultati di qualsiasi compito o richiesta riflettano quanto appreso nell'addestramento su quel problema specifico o simile, o se invece il sistema abbia acquisito la capacità autonoma di risolvere i problemi posti. Altre preoccupazioni riguardo ai dati utilizzati per addestrare gli LMM, compresa la conformità alle leggi sulla protezione dei dati, sono discusse di seguito.
Né la singola persona né i governi erano preparati per il rilascio degli LMM. La persona media non è stata educata all'utilizzo degli LMM e potrebbe non capire che le risposte non sono sempre accurate o affidabili, anche se un chatbot alimentato da LMM crea tale impressione. Uno studio ha segnalato che, sebbene un modello linguistico di grandi dimensioni come GPT-3 possa "produrre informazioni accurate più facili da comprendere rispetto a un essere umano", può anche generare una "disinformazione più convincente", e una persona non è in grado di distinguere i contenuti generati da un LMM da quelli generati da un essere umano [16].

Anche i governi si sono mostrati in gran parte impreparati. Le normative e le leggi per governare l'uso dell'IA possono non essere adatte ad affrontare le sfide e le opportunità associate agli LMM. L'Unione Europea, che ha raggiunto un accordo per promulgare l'*Artificial Intelligence Act* a livello europeo, ha dovuto poi revisionare il proprio quadro legislativo nelle fasi finali della stesura per tener conto degli LMM [17]. Altri governi nazionali stanno sviluppando rapidamente nuove leggi o regolamenti [18] o hanno istituito divieti temporanei (alcuni dei quali sono già stati revocati) [19]. Si prevede che le aziende produttrici rilasceranno successivamente LMM sempre più potenti e capaci nei prossimi

mesi, e ciò significa nuovi benefici ma anche nuove sfide normative. In questo ambiente così dinamico, questa guida, che si basa su precedenti orientamenti, inclusi quelli sull'etica, fornisce suggerimenti e raccomandazioni per l'uso degli LMM nell'assistenza sanitaria e nella medicina.

1.2 Linee guida OMS sull'etica e la governance dell'IA in sanità

Le linee guida originali dell'OMS sull'etica e la governance dell'IA in sanità [1] hanno esaminato diversi approcci al machine learning e alle varie applicazioni dell'IA nell'assistenza sanitaria, ma non hanno esaminato nello specifico l'IA generativa o gli LMM. Durante lo sviluppo di tali linee guida e al momento della loro pubblicazione nel 2021, non c'era alcuna avvisaglia che l'IA generativa e gli LMM sarebbero stati ampiamente disponibili in tempi così rapidi e che sarebbero stati applicati nella clinica, nella ricerca medica e nella sanità pubblica.

Nonostante ciò, le sfide etiche identificate nelle linee guida e i principi e le raccomandazioni etiche fondamentali (vedi **Box 1**) rimangono validi sia per valutare sia per utilizzare in modo efficace e sicuro gli LMM, anche se ulteriori difetti di governance e nuove sfide continuano a emergere in merito a questa nuova tecnologia. Le sfide, i principi e le raccomandazioni sono alla base dell'approccio agli LMM presentato in queste linee guida.

Box 1. *Sintesi del consensus OMS sui principi etici per l'utilizzo dell'IA in sanità*

Tutelare l'autonomia: gli esseri umani devono mantenere il controllo dei sistemi sanitari e delle decisioni mediche. I professionisti sanitari dispongono delle informazioni necessarie per utilizzare in modo sicuro ed efficace i sistemi di IA. Le persone comprendono il ruolo che i sistemi di IA svolgono nella loro cura. La privacy dei dati e la riservatezza sono protette da un consenso informato valido rispettando le norme giuridiche per la protezione dei dati.

Promuovere il benessere e la sicurezza delle persone e l'interesse pubblico: i progettisti di IA soddisfano i requisiti normativi di sicurezza, accuratezza ed efficacia per utilizzi o indicazioni ben definiti. Dovrebbero essere disponibili misure di controllo della qualità nella pratica e di miglioramento della qualità nell'uso dell'IA nel tempo. L'IA non è utilizzata se comporta danni mentali o fisici che potrebbero essere evitati mediante l'uso di una pratica o di un approccio alternativo.

Garantire trasparenza, "spiegabilità" e comprensibilità: le tecnologie dell'IA devono essere intelligibili o comprensibili per sviluppatori, professionisti sanitari, pazienti, utenti e regolatori. Prima della progettazione o della distribuzione dell'IA vengono pubblicate o documentate le informazioni sufficienti; l'informazione agevola una consultazione pubblica e un dibattito significativi su come l'IA è stata progettata e come dovrebbe o non dovrebbe essere utilizzata. L'IA può essere spiegata tenendo presente la capacità di comprensione di coloro a cui viene spiegata.

Favorire responsabilità e rendicontabilità (accountability) al fine di garantire che l'IA sia utilizzata in condizioni appropriate e da persone adeguatamente formate. Pazienti e medici valutano lo sviluppo e l'implementazione dell'IA. A monte e a valle dell'algoritmo sono applicati principi regolatori definendo dei momenti di supervisione umana. Sono disponibili meccanismi appropriati per la richiesta e per il risarcimento di persone e gruppi che sono stati danneggiati da decisioni basate sull'IA.

Garantire inclusività ed equità: l'IA è progettata e condivisa per incoraggiare l'uso e l'accesso il più ampio, appropriato ed equo possibile, indipendentemente dall'età, dal sesso, dall'identità di genere, dal reddito, dalla razza, dall'etnia, dall'orientamento sessuale, dalle capacità o da altre caratteristiche. L'IA è disponibile per l'uso non solo nei contesti ad alto reddito, ma anche nei Paesi a basso e medio reddito. L'IA non ricorre a pregiudizi a svantaggio di gruppi identificabili. L'IA minimizza le inevitabili disparità di potere. L'IA è monitorata e valutata per identificare eventuali effetti discriminanti su specifici gruppi di persone.

Promuovere un'IA responsiva e sostenibile: le tecnologie di IA sono coerenti con la promozione più ampia della sostenibilità dei sistemi sanitari, dell'ambiente e dei luoghi di lavoro.

→ Torna all’indice

I. Applicazioni, sfide e rischi degli LMM

2. Applicazioni e sfide connesse all'utilizzo degli LMM in ambito sanitario

L'ambito d'uso dell'IA in sanità comprende la diagnosi, l'assistenza clinica, la ricerca, lo sviluppo di farmaci, l'amministrazione sanitaria, la sanità pubblica e la sorveglianza. Molti dei modi di utilizzo degli LMM non sono innovativi dell'IA, ma in ogni caso clinici, pazienti, cittadini ed esperti in ambito sanitario fanno un uso diverso degli LMM. Questa sezione discute dei potenziali utilizzi degli LMM in ambito sanitario con le previsioni per quanto riguarda potenziali sfide e rischi associati al loro utilizzo. Molti modi d'uso sono ancora in fase sperimentale e potrebbero alla fine non essere in grado di offrire i benefici previsti.

2.1 Diagnosi e assistenza clinica

L'IA è già utilizzata in ambito di diagnosi e di assistenza clinica, per esempio come supporto nella diagnostica per immagini, nelle malattie infettive e in oncologia. Si è ipotizzato che i clinici possano usare l'IA per integrare la cartella clinica dei pazienti durante la consultazione, per i pazienti a rischio e come un aiuto per prendere decisioni di trattamento complesse e per identificare gli errori clinici [1]. Gli LMM potrebbero estendere l'uso di sistemi basati sull'IA nella diagnosi e nell'assistenza clinica – sia virtuale sia in presenza – tanto che, secondo alcuni esperti, gli LMM "saranno più importanti per i medici di quanto non fosse lo stetoscopio in passato" [20]. Molti LMM hanno superato l'esame di licenza per esercitare la medicina negli Stati Uniti, anche se superare un esame di medicina scritto grazie all'enorme mole di conoscenze mediche non è la stessa cosa che fornire poi servizi clinici sicuri ed efficaci [21]. Gli LMM non sono riusciti inoltre a superare esami con materiale che non fosse stato precedentemente pubblicato online o facilmente risolvibili da bambini [22]. Uno studio sulla conoscenza clinica di un modello di linguaggio di grandi dimensioni ha concluso che "la transizione da un LLM usato per rispondere a domande di medicina a uno strumento che possa essere utilizzato da professionisti sanitari, amministratori e consumatori necessita di ulteriori ricerche per assicurare la sicurezza, l'affidabilità, l'efficacia e la privacy della tecnologia" [23].

La diagnosi è vista come un'area particolarmente promettente, perché gli LMM potrebbero essere d'aiuto nel diagnosticare malattie rare o "presentazioni insolite" in casi complessi [24]. I medici già utilizzano motori di ricerca in Internet, risorse online e altri supporti alla diagnosi, e gli LMM sarebbero un ulteriore strumento per la diagnosi. Gli LMM potrebbero essere usati anche per diagnosi di routine, per fornire ai medici un'opinione in più per essere sicuri che diagnosi ovvie non vengano ignorate. Tutto ciò può essere fatto in tempi molto rapidi e ciò è in parte dovuto al fatto che gli LMM posso-

no esaminare l'intero fascicolo sanitario di un paziente molto più velocemente di un medico [24]. Alcuni LMM molto noti che sono ora adoperati in programmi pilota come supporto clinico non sono comunque stati allenati su fascicoli sanitari elettronici, né su dati di medicina o altri dati sanitari, anche se i loro dataset includono tali informazioni. Per esempio, in molti sistemi sanitari degli Stati Uniti un LMM fornito da un'azienda tecnologica viene valutato in via sperimentale per leggere messaggi inviati dai pazienti e abbozzare le risposte dei medici al fine di ridurre il tempo che il personale spende nel rispondere alle domande dei pazienti. Questa pratica è volta alla riduzione del "burnout" dei professionisti sanitari che inviano migliaia di messaggi al giorno, dandogli la possibilità di concentrarsi sulla loro attività clinica ("liberazione dalla tastiera") [25]. Quando il messaggio di un paziente viene ricevuto, l'LMM ipotizza una bozza di risposta basata sia sull'informazione del paziente sia sul suo fascicolo sanitario elettronico. Nonostante l'IA venga utilizzata solo per alcuni tipi di domande, le risposte fornite richiedono comunque un ulteriore sforzo di rifinitura da parte dei medici non indifferente [25].

Cionondimeno, uno studio statunitense ha mostrato come un chatbot basato su ChatGPT abbia performato meglio di medici qualificati nel rispondere a domande di salute poste in un forum online. Delle 195 domande selezionate, i revisori hanno valutato come migliori le risposte del chatbot rispetto a quelle dei medici nell'80% circa dei casi [26]. I chatbot potrebbero essere utili anche per rispondere a domande standard di "consultazioni informali" e per fornire informazioni e risposte sulla presentazione iniziale del paziente o per sintetizzare i risultati degli esami di laboratorio [27].
Imprese e università stanno inoltre sviluppando LMM allenati con dati medici e sanitari o con fascicoli sanitari elettronici, inclusi LMM basati su piccoli dataset. Per esempio, un LMM è stato allenato su un dataset di circa 30.000 report di casi per imparare a cogliere la relazione tra sintomi e malattia per fornire aiuto nel procedimento diagnostico [28]. Un altro LMM è stato allenato su un dataset di oltre 100.000 radiografie del torace per identificare anomalie ed eventualmente fornire informazioni valide o identificare condizioni di malattia [29]. Molti LMM, pubblicamente valutati, sono stati allenati con un algoritmo su milioni di fascicoli sanitari elettronici e altre fonti di conoscenza medica generale e specializzata. L'approccio ha migliorato la capacità dell'algoritmo di processare diverse forme di informazione medica scritta e formulare risposte ("rispondere a domande mediche") [30].

Molte grandi aziende tecnologiche stanno adattando i loro LMM realizzati per scopi generali verso modelli che potrebbero fornire assistenza per la diagnosi e per la cura. Un'azienda sta sviluppando Med-PaLM 2, che dovrebbe poter rispondere a domande e riassumere informazioni a partire da testi medici, e si sta ora evolvendo per descrivere immagini (come radiografie), per scrivere report e rispondere a domande sul follow-up, per facilitare ulteriori domande da parte dei medici, una funzione che potrebbe ridurre l'eventuale disaccordo di opinione tra un professionista sanitario e un computer [31]. L'obiettivo a lungo termine è sviluppare una "intelligenza artificiale medica generalista", che permetterebbe ai professionisti sanitari di dialogare in maniera flessibile con un LMM per generare risposte in base a domande personalizzate in ambito clinico. Un utilizzatore potrebbe adattare un modello di IA medica generalista a un nuovo scopo, descrivendo ciò che si vuole ottenere in linguaggio comune, senza dover allenare da capo l'LMM né doverlo allenare ad accettare differenti tipi di dati non strutturati per generare una risposta [5].

Rischi dell'utilizzo degli LMM per la diagnosi e l'assistenza clinica

L'idea di includere gli LMM nell'assistenza clinica si accompagna a rischi significativi associati al loro utilizzo, molti dei quali sono antecedenti agli LMM. Sono stati individuati cinque rischi principali derivanti dall'utilizzo di un LMM nella diagnosi e nell'assistenza clinica.

A. **Risposte imprecise, incomplete, parziali o false:** una delle preoccupazioni nell'utilizzo degli LMM è la propensione dei chatbot a produrre risposte non completamente corrette o del tutto false a partire da dati o informazioni (come le citazioni) "inventate" dall'LMM [32] e risposte che siano viziate da bias che derivano dai difetti in fase di addestramento [33]. Gli LMM potrebbero inoltre contribuire a bias di contestualizzazione quando le assunzioni su dove sia utilizzata una tecnologia IA porti a fare raccomandazioni in realtà sviluppate per l'utilizzo in un altro contesto [1]. Per esempio, nei dati usati per il training c'è una sotto rappresentazione dei dati e delle prospettive dei Paesi a reddito medio-basso. Per questo motivo se a un LMM fosse chiesto di definire lo schema di un trattamento per una malattia al fine di guidare le scelte di un ministero della salute di un Paese a basso reddito, l'LMM potrebbe descrivere un approccio appropriato solo in un Paese ad alto reddito [34]. Inoltre, un LMM potrebbe fornire una risposta incompleta, non fornire affatto una risposta o fornire una risposta che non tenga in considerazione il cambiamento delle circostanze nell'ambito in cui viene utilizzata.

Le risposte false, note colloquialmente come "allucinazioni", sono indistinguibili dalle risposte fattuali e corrette generate da un LMM, perché gli LMM, nemmeno quelli allenati per rinforzo tramite feedback umano, non sono allenati alla produzione di fatti, ma alla produzione di informazioni che sembrano fattuali. Uno studio ha mostrato come gli LLM quando venga loro fornito un semplice insieme di fatti da riassumere hanno allucinazioni in almeno il 3% dei casi, fino al 27% [35]. Attualmente gli LMM dipendono anche dal prompt engineering umano – cioè dalla richiesta che viene scritta – in cui l'input viene ottimizzato per comunicare in maniera efficace con l'LMM [36]. Dunque gli LMM, anche se allenati in maniera specifica su dati medici e informazioni sanitarie, potrebbero comunque non produrre risposte corrette. Per alcune diagnosi fatte dall'LMM potrebbe non esserci alcun test di conferma, né altri modi per verificarne la correttezza [24].
In medicina e in altre aree decisionali della sanità pubblica, l'uso degli LMM, anche qualora fossero fattualmente corretti nella maggior parte dei casi, potrebbe non essere così accurato da giustificare il costo del loro sviluppo o della loro implementazione sicura ed efficace nei sistemi sanitari.

B. **Qualità dei dati e bias sui dati:** uno dei motivi per cui gli LMM producono risposte parziali o imprecise è la bassa qualità dei dati. Molti degli LMM attualmente disponibili per uso pubblico sono stati allenati su grandi dataset, come su Internet, e ciò potrebbe favorire la diffusione di una cattiva informazione e di bias. La maggior parte dei dati medici e sanitari hanno pure un bias di selezione, per esempio per razza, etnia, discendenza, sesso, identità di genere o età. Gli LMM allenati su dati sanitari spesso fanno propri questi bias, visto che la maggior parte dei dati sanitari viene raccolta nei Paesi ad alto reddito. Per esempio, i dati genetici sono raccolti molto

più spesso in persone di discendenza europea [1]. Gli LMM sono inoltre spesso allenati su record sanitari elettronici, che sono a loro volta pieni di errori e di informazioni imprecise [24] oppure si basano su informazioni raccolte durante una visita medica, che possono essere poco accurate, influenzando così la risposta fornita da un LMM [25]. I problemi di qualità dei dati e di bias riguardano tutti i modelli di IA, compresi gli LMM [1].

L'azienda tecnologica di GPT-4 dichiara nella "carta di sistema": "Abbiamo scoperto che 'GPT-4-early' e 'GPT-4-launch' hanno gli stessi difetti di modelli di linguaggio precedenti, come la produzione di contenuti parziali e inaffidabili" [37]. I limiti degli LMM possono essere legati alla data di termine di input dei dati su cui sono stati addestrati, nonostante alcuni LMM possano ora accedere a informazioni aggiornate direttamente da Internet. Per esempio, ChatGPT-4 è stato allenato su dati fino a settembre 2021 [38], e via via aggiornato ma può cercare su Internet informazioni più recenti [39]. Ciò potrebbe portare alla generazione di ulteriori informazioni false o imprecise, mentre in passato l'indicazione di una data di termine dell'aggiornamento dei dati garantiva dall'introduzione successiva di nuovi materiali falsi pubblicati [39]. In medicina è indispensabile che le informazioni siano aggiornate e molto accurate per mantenere elevati gli standard di assistenza e per la comprensione di alcune malattie.

C. **Bias dell'automazione:** il problema della generazione di risposte false, imprecise o influenzate da bias da parte degli LMM è intensificato dal fatto che, come con altre forme di IA, gli LMM inducono un bias di automazione negli esperti, nei professionisti sanitari e nei pazienti (vedi oltre). Grazie al bias di automazione, un medico potrebbe non rilevare errori che sarebbero stati individuati da un essere umano [1]. C'è anche il timore che medici e operatori sanitari possano utilizzare gli LMM per prendere decisioni nelle quali occorre considerare anche aspetti etici o morali [20]. Gli LMM come ChatGPT possono essere molto incoerenti come "consiglieri morali", nonostante, come indicano esperimenti recenti, siano in grado di influenzare il giudizio morale degli utilizzatori, anche se questi sono consapevoli di ricevere consigli da un chatbot [40]. L'utilizzo di LMM per giudizi morali potrebbe portare a una deresponsabilizzazione riguardo alla morale, per cui i medici diverrebbero incapaci di giudicare o di prendere decisioni complesse [20].

D. **Degradazione delle competenze:** c'è il rischio a lungo termine che l'uso dell'IA nella pratica medica possa degradare o erodere le competenze dei clinici come professionisti, a causa di un progressivo trasferimento delle loro responsabilità ai computer. La perdita di competenze potrebbe portare a medici incapaci di prevalere o di criticare con convinzione la decisione di un algoritmo o, nell'eventualità di un errore di sistema o di una breccia nel sistema di sicurezza, a medici non in grado di portare a termine eventuali attività mediche o procedure senza l'ausilio dell'IA [1].

E. **Consenso informato:** l'uso crescente degli LMM, di persona ma soprattutto virtualmente, deve garantire che il paziente sia consapevole che gli sta rispondendo una IA e che la risposta potrebbe sembrare quella di un medico. Eppure, dove e quando gli LMM o altre forme di IA vengono utilizzate regolarmente nella pratica medica, i pazienti o i loro caregiver, pur non sentendosi

a proprio agio o non volendo fare affidamento su una tecnologia IA, potrebbero non avere la scelta di negare il consenso del suo utilizzo. Ciò è vero specialmente se altre opzioni (non basate sull'IA) non sono facilmente disponibili o se il medico, che ha affidato la responsabilità di certe funzioni al computer, non è in grado di fornire assistenza senza l'utilizzo dell'IA.

2.2 Applicazioni centrate sul paziente

L'IA sta iniziando a cambiare il modo in cui i pazienti controllano le proprie condizioni mediche. I pazienti già si assumono responsabilità significative riguardo alla propria cura, incluso prendere farmaci, migliorare la propria nutrizione e dieta, fare attività fisica, trattare ferite e praticare iniezioni. Gli strumenti di IA sono stati progettati per migliorare la cura di sé grazie anche all'uso di chatbot, sistemi di monitoraggio della salute e di predizione di rischi e sistemi creati per le persone con disabilità [1]. Gli LMM potrebbero accelerare la tendenza all'uso di IA da parte di pazienti e cittadini per scopi medici. Da vent'anni le persone usano Internet per cercare informazioni mediche. Gli LMM potrebbero avere un ruolo centrale nel fornire informazioni ai pazienti e ai cittadini, tramite la loro integrazione con le ricerche in Internet. Chatbot alimentati da un LMM potrebbero sostituire Internet per la ricerca di informazioni [41], per avere un'autodiagnosi e per ottenere informazioni prima di andare dal medico.

Chatbot alimentati da LMM, con dati di tipologie sempre più varie, potrebbero servire come assistenti sanitari virtuali altamente personalizzati su temi generali. Secondo uno studio, "gli assistenti sanitari virtuali possono fare leva sulle singole persone per promuovere modifiche comportamentali, rispondere a domande in ambito sanitario, valutare sintomi, o comunicare con operatori sanitari quando necessario" [3]. Nello specifico, chatbot alimentati da LMM potrebbero fornire trattamenti, per esempio, per la salute mentale [2].

Una terza applicazione degli LMM centrati sul paziente potrebbe essere volta all'identificazione di sperimentazioni cliniche o all'inclusione in uno studio [28]. Mentre programmi basati sull'IA già servono a pazienti e ricercatori per metterli in contatto [42], gli LMM potrebbero essere usati in maniera analoga usando dati medici rilevanti del paziente [28]. Tale uso dell'IA potrebbe sia ridurre il costo del reclutamento nello studio sia rendere più rapido ed efficiente l'arruolamento, dando al contempo più opportunità alle persone di trovare sperimentazioni e trattamenti appropriati che sono difficili da identificare attraverso altri canali [42].

Rischi e sfide

La facilità con cui gli LMM possono essere utilizzati dalle persone può comportare rischi significativi, tra cui quelli elencati di seguito.

- **Affermazioni inaccurate, non complete o false:** come per l'uso degli LMM da parte dei medici e degli altri professionisti sanitari, il loro utilizzo da parte dei pazienti e dei non addetti ai lavori è associato al rischio di avere informazioni false, distorte, incomplete o imprecise anche da parte di programmi di IA specializzati nel fornire informazioni mediche. I rischi sono maggiori quando

questi programmi vengono utilizzati da persone prive di competenza medica o che non hanno le basi per valutare criticamente la risposta avuta, o che non hanno accesso a un'altra fonte di informazioni, oppure quando questi programmi vengono usati dai bambini (vedi **Box 2**). Sebbene le persone da alcuni decenni cerchino in Internet le informazioni mediche, un LMM può fornire risposte che sembrano apparentemente corrette, facendo riferimento solo ad altri LMM (che comportano gli stessi rischi) per avere un rapido confronto.

- **Manipolazione:** molte applicazioni di chatbot usate con gli LMM hanno approcci particolari al dialogo, che si prevede diventino più persuasivi e più coinvolgenti [46] e i chatbot potrebbero anche essere in grado di adattare i modelli di conversazione a ciascun utente [41]. I chatbot possono fornire risposte a domande o impegnarsi in conversazioni per persuadere le persone a compiere azioni contrarie al loro interesse o benessere [12]. Diversi esperti hanno chiesto un intervento urgente per gestire le potenziali conseguenze negative dei chatbot, notando che potrebbero diventare "emotivamente manipolativi" [47,48]. Un caso molto noto è quello in Belgio di una persona affetta da disturbo d'ansia generalizzata, che si è suicidata dopo 6 settimane di conversazioni intensive con un chatbot [49].
- **Privacy:** l'uso degli LMM da parte dei pazienti e dei non "addetti ai lavori" potrebbe non essere riservato e potrebbe non rispettare la riservatezza dei dati personali e sanitari che vengono condivisi. Coloro che hanno utilizzato gli LMM per altri scopi (non sanitari) hanno tendenzialmente condiviso informazioni sensibili, come informazioni aziendali proprietarie [50]. I dati condivisi su un LMM non necessariamente vengono cancellati, poiché le aziende tecnologiche potrebbero utilizzarli per migliorare i propri modelli di IA [50], anche qualora non ci sia una base legale che

Box 2. *Considerazioni etiche circa l'uso degli LMM da parte dei bambini*

Mentre di recente sono state pubblicate linee guida generali per la sicurezza e per l'utilizzo etico dei dati pediatrici nell'IA e nel machine learning (ACCEPT-AI) [43], occorre fare considerazioni mirate sul potenziale impatto dell'uso degli LMM da parte dei bambini.

L'ampia disponibilità degli LMM fornisce l'accesso a utilizzatori di varie età, ma ci sono pochi dati su come i bambini interagiscano o utilizzino gli LMM. Mentre il potenziale di opportunità e i problemi connessi agli LMM sono stati discussi in contesti educativi più generali [44], non è chiaro come l'interazione con un LMM da parte dei bambini possa influenzare il loro benessere mentale o fisico. L'utilizzo degli LMM da parte dei bambini deve essere monitorato nel tempo per comprenderne i benefici e gli eventuali problemi.

Le leggi e i regolamenti sul consenso pediatrico, sull'assenso e sul coinvolgimento legale dei genitori differiscono tra Paesi. Per questo motivo la mancanza di regole specifiche e coerenti tra loro, unificate e globali per l'uso degli LMM da parte dei bambini potrebbe far emergere problemi non identificati e non controllati. Nello specifico, non è chiaro con quale accuratezza gli LMM distinguano tra dati pediatrici e dati della popolazione adulta. Alcuni studi hanno mostrato che la generalizzazione di dati ottenuti da adulti alla popolazione pediatrica negli LMM può essere limitata. I dati pediatrici dovrebbero dunque essere tenuti separati nei dataset di allenamento e di test [45].

Gli sviluppatori dovrebbero includere informazioni demografiche sui dati di allenamento che includano l'età e vanno incoraggiati a fornire descrittori chiari delle popolazioni di riferimento, compreso l'intervallo di età, per un'interazione appropriata e sicura con gli LMM. Quando legalmente possibile, gli LMM dovrebbero essere migliorati includendo interazioni appropriate e feedback da parte degli utenti più giovani.

lo consenta o sebbene i dati possano essere eventualmente rimossi dai server aziendali [51]. Un problema correlato è la condivisione di informazioni su un LMM con altri utenti dello stesso LMM sia a seguito della richiesta specifica da parte di un altro utente dell'LMM di fornire tali informazioni [52], sia per una divulgazione fatta per errore o comunque non autorizzata della cronologia delle chat di interrogazione dell'LMM di altre persone (anche se non il contenuto della conversazione) [53]. Pertanto, se le informazioni mediche identificabili di una persona vengono fornite a un LMM, potrebbero essere divulgate a terze parti [54].

- **Degrado delle interazioni tra medici, cittadini e pazienti:** l'uso degli LMM da parte dei pazienti o di chi li assiste potrebbe cambiare radicalmente il rapporto medico-paziente. L'aumento delle ricerche su Internet da parte dei pazienti negli ultimi vent'anni ha già modificato questo rapporto, in quanto i pazienti possono utilizzare le informazioni che trovano per contestare o chiedere maggiori informazioni al loro medico. Se da un lato un LMM potrebbe migliorare questo dialogo, dall'altro un paziente o un caregiver potrebbe decidere di affidarsi completamente a un LMM per la prognosi e il trattamento, riducendo o rendendo superfluo il ricorso al giudizio e al supporto di un medico. Una preoccupazione correlata è che, se una tecnologia di IA riduce il contatto tra il professionista e il paziente, ridurrebbe anche le opportunità per i medici di promuovere la salute e potrebbe minare l'assistenza sanitaria in generale, come le interazioni interpersonali fondamentali quando le persone sono più vulnerabili [1]. C'è quindi la preoccupazione che l'assistenza medica possa essere "disumanizzata" dall'IA.
- **Ingiustizia epistemica:** un'altra possibile conseguenza della sostituzione del giudizio di un professionista sanitario con quello di un LMM è l'introduzione di un'ingiustizia epistemica per il paziente. L'ingiustizia epistemica è un "torto fatto a qualcuno nella sua qualità di portatore di conoscenza", come un paziente in un sistema sanitario [55].
  Una forma di ingiustizia epistemica, l'ingiustizia ermeneutica, si verifica quando c'è una lacuna nella comprensione e nella conoscenza condivisa (le cosiddette "risorse interpretative collettive") che pone alcune persone in una posizione di svantaggio rispetto alla loro esperienza vissuta, all'esperienza sociale o, nel caso della salute, alla comprensione della propria condizione fisica o mentale [55]. Gli LMM, anche se addestrati su grandi quantità di dati, hanno dei limiti rispetto a ciò che possono riconoscere e a cui possono rispondere e ai concetti e alle nozioni che non rientrano nel loro vocabolario. Se l'esperienza di un paziente non viene compresa o riconosciuta da un LMM in un ambito clinico, ciò può impedire un'assistenza appropriata da parte di un medico, con il rischio di danni per il paziente. Tale rischio è più elevato nei gruppi vulnerabili, che sono già trascurati e sotto rappresentati nei dati [55], come le persone con disabilità (vedi **Box 3** a pagina 36).
- **Fornitura di assistenza sanitaria al di fuori del sistema sanitario:** le applicazioni di IA per la salute non vengono utilizzate esclusivamente all'interno dei sistemi sanitari o nell'assistenza domiciliare, poiché le tecnologie di IA per la salute possono essere facilmente acquisite e utilizzate da enti non appartenenti al sistema sanitario o semplicemente introdotte da un'azienda, per esempio quelle stesse che offrono gli LMM per un uso pubblico. Ciò solleva la questione se

tali tecnologie debbano essere regolamentate come applicazioni mediche, che richiedono un maggiore controllo normativo, oppure come "applicazioni per il benessere", che invece sono sottoposte a meno regole. Attualmente, tali tecnologie rientrano in una zona grigia tra le due categorie.

Gli LMM regolamentati in modo non stringente possono essere un rischio se vengono utilizzati da un paziente senza che vi sia alcuna tutela normativa. Ciò include l'uso di un LMM per avere un parere medico o per farsi una autodiagnosi. Il timore è che i pazienti possano ricevere consigli non corretti o fuorvianti (vedi sopra) e che la loro sicurezza possa essere compromessa se le persone non entrano in contatto con i servizi sanitari, compresa la mancanza di cure di supporto psicologico per le persone con ideazione suicidaria che utilizzano un chatbot di IA, pure se il chatbot non è "manipolativo". Anche nel caso in cui le informazioni derivanti dal chatbot di IA risultassero corrette, le persone senza una formazione medica che utilizzino tali informazioni per una autodiagnosi potrebbero interpretarle o utilizzarle in modo errato. Poiché tali applicazioni, compresi gli LMM, continuano a proliferare e non sono necessariamente registrate come applicazioni mediche, la qualità complessiva dell'assistenza sanitaria potrebbe essere compromessa. Ciò potrebbe ulteriormente esacerbare le diseguaglianze nell'accesso a una assistenza sanitaria di buona qualità, specialmente perché le persone prive di altre opzioni potrebbero ricorrere a tali applicazioni [1].

2.3 Compiti amministrativi e funzioni di trascrizione

Gli LMM stanno iniziando a essere usati per assistere gli operatori negli aspetti amministrativi e finanziari della pratica medica. Sempre più spesso i medici e gli altri professionisti sanitari in molti

Box 3. *Considerazioni etiche associate con gli LMM e loro influenza sulle persone con disabilità*

Poiché in passato le persone con disabilità sono state spesso escluse dai luoghi di lavoro, dalla scuola e dal ricevere un adeguato supporto medico [56], i dataset utilizzati per l'addestramento di sistemi di IA non le contemplano.

I sistemi potrebbero discriminare persone con asimmetrie facciali, diversi modi di gesticolare, stili di comunicazione, di comportamento o schemi motori. Il gruppo che ne è più affetto sono le persone con disabilità, deficit cognitivi o sensoriali o disturbi dello spettro autistico [57].

Tale bias ed esclusione può applicarsi all'IA generativa. Per esempio, gli LMM potrebbero introdurre una connotazione o un sentimento negativi in frasi o parole associate con la disabilità nella descrizione o nell'anamnesi di un paziente [58]. I chatbot potrebbero identificare una persona con disabilità come "non vivente", "non umano" o "emotivamente piatto" a causa di atteggiamenti o pattern di azioni. I sistemi di riconoscimento vocale potrebbero essere meno accurati nelle persone con difetti linguistici, portando a interpretazioni errate.

Durante lo sviluppo dell'IA bisogna considerare e risolvere i problemi e i bias relativi alla disabilità: l'inclusione di persone con disabilità nello sviluppo e nella progettazione dei sistemi di IA aiuterebbe a valutare i bias legati alle disabilità e le performance di un sistema di IA, assicurandosi che le leggi emanate per proteggere e promuovere i diritti delle persone con disabilità si occupino delle problematiche associate con la tecnologia IA, promuovendo al contempo leggi e politiche che regolino l'IA per le probabili sfide e barriere affrontate dalle persone con disabilità, con l'aumento dell'uso di sistemi basati sull'IA. Leggi specifiche per l'IA dovrebbero includere categorizzazioni di specifiche disabilità, incluso come singole condizioni siano influenzate dai sistemi di IA.

contesti sono sottoposti a una crescente mole di lavoro cartaceo per i numerosi obblighi di registrazione delle informazioni e dei dati dei pazienti nelle cartelle cliniche elettroniche, di fatturazione nei sistemi sanitari privati, assicurativi o pubblici e di altri compiti amministrativi. Sebbene molti di questi obblighi, come la compilazione delle cartelle cliniche elettroniche, fossero stati pensati per liberare più tempo a professionisti sanitari per svolgere le attività cliniche, la maggior parte di essi è oggi una delle principali cause di burn-out di medici e altri professionisti sanitari [59]. Secondo uno studio pubblicato in letteratura, la documentazione prende da un quarto alla metà del tempo di un medico e un quinto di quello di un infermiere [59].
Gli LMM sono stati identificati come il possibile mezzo per restituire ai professionisti sanitari il bene più prezioso, il tempo, sia per ridurre il burn-out, sia per dedicare più tempo al singolo paziente, sia per vedere più pazienti. Un medico che ha utilizzato un software che inglobava un LMM per documentare le visite ai pazienti ha dichiarato: "l'intelligenza artificiale mi ha permesso, come medico, di essere disponibile al 100% per i miei pazienti" e ha aggiunto che il software gli ha consentito di ottimizzare la sua gestione del tempo, permettendogli di risparmiare fino a 2 ore al giorno [60].

Alcuni esempi di usi attuali e prevedibili degli LMM in questo ambito sono:

- **una migliore comunicazione in supporto alla traduzione o al miglioramento della comunicazione medico-paziente, semplificando il gergo medico e rendendo la comunicazione più "a misura di paziente" [34]**
- **completamento delle informazioni mancanti nelle cartelle cliniche elettroniche [61]**
- **scrittura delle note cliniche dopo ogni visita del paziente (virtuale o in presenza) [62]**
- **si prevede che l'uso degli LMM consentirà anche la stesura automatizzata di prescrizioni, appuntamenti, fatturazioni, programmazione di esami, pre-autorizzazioni da parte delle compagnie assicurative, esiti di esami e lettere di dimissione [5]. Con lo sviluppo, LMM più sofisticati potrebbero essere impiegati per la stesura di cartelle cliniche ancora più complesse, per esempio per i radiologi, "redigendo automaticamente referti radiologici che descrivono sia le anomalie sia i reperti normali, tenendo conto dell'anamnesi del paziente... questi modelli possono fornire un'ulteriore aiuto ai medici abbinando ai referti testuali una visualizzazione interattiva, per esempio evidenziando la regione anatomica descritta in ogni passaggio del testo" [5].**

Rischi e sfide

Come per altri usi degli LMM, potrebbero verificarsi gravi errori, dovuti a imprecisione, difetti nella trascrizione, traduzione o semplificazione o ancora allucinazioni. È quindi importante che la maggior parte delle funzioni amministrative e di trascrizione non vengano completamente automatizzate. Anche se la supervisione e la revisione richiedono il tempo dedicato da un professionista sanitario, è probabile che l'impiego di un LMM sia comunque meno oneroso di quanto accada ora. Un altro limite è che gli LMM possono essere incoerenti: lievi modifiche a un prompt o a una domanda possono generare una cartella clinica elettronica completamente diversa, anche se si prevede che queste incoerenze diminuiranno nel tempo [63].

2.4 Formazione medica e infermieristica

Si prevede che gli LMM svolgeranno un ruolo attivo anche nella formazione medica e infermieristica. Potrebbero essere utilizzati per creare testi dinamici che, rispetto ai testi generici, si adattino alle esigenze e alle richieste specifiche di uno studente [63]. Gli LMM integrati nei chatbot possono permettere conversazioni simulate per migliorare la comunicazione medico-paziente e la risoluzione dei problemi, tra i quali la pratica del colloquio medico, il ragionamento diagnostico e la spiegazione delle opzioni terapeutiche. Un chatbot potrebbe anche essere personalizzato per sottoporre a uno studente vari pazienti virtuali, compresi quelli con disabilità o condizioni mediche rare. Gli LMM potrebbero anche fornire istruzioni seguendo le quali uno studente di medicina pone domande e riceve risposte accompagnate da un ragionamento, attraverso una catena di pensieri che includono i processi fisiologici e biologici [63].

Rischi e sfide

Sebbene l'uso dell'IA possa migliorare o affinare la formazione e le competenze di un professionista sanitario, potrebbe anche comportare il rischio che i professionisti sospendano il loro giudizio (o quello di un collega) a favore di quello di un computer. Se un LMM fornisce informazioni o risposte non corrette o inventa una risposta, potrebbe influire sulla qualità della formazione medica.
Un'ulteriore preoccupazione è che l'uso degli LMM nella formazione o per semplificare le funzioni amministrative e di trascrizione potrebbe comportare un onere aggiuntivo per gli operatori sanitari che non sono ancora alfabetizzati digitalmente e dovranno quindi sviluppare nuove competenze nell'uso delle tecnologie supportate dall'IA nella pratica quotidiana [1]. Si prevede che le nuove funzionalità degli LMM richiedano ai professionisti sanitari una continua riqualificazione e adattamento [1]. Gli sviluppatori potrebbero introdurre tecnologie supportate dall'IA con interfacce di comunicazione facilmente utilizzabili dai non addetti ai lavori, come il linguaggio naturale o la visione.

2.5 Ricerca medica e scientifica e sviluppo di nuovi farmaci

L'IA è già utilizzata nella ricerca scientifica e clinica e nello sviluppo di farmaci. Con l'IA è possibile analizzare le cartelle cliniche elettroniche per identificare modelli di pratica clinica e svilupparne di nuovi. L'apprendimento automatico viene utilizzato anche nell'ambito della genomica, per esempio per migliorare la comprensione di una malattia e identificare nuovi biomarcatori [1].
L'IA viene utilizzata in quasi tutte le fasi del ciclo di sviluppo dei farmaci, per ottimizzare lo screening delle molecole candidate, per prevedere la forma tridimensionale di una proteina (il "problema del ripiegamento delle proteine") [1], per prevedere la tossicità e l'efficacia dei composti in fase di sviluppo preclinico e per migliorare il reclutamento, l'arruolamento e il monitoraggio dei pazienti durante gli studi clinici.[1]
Gli LMM stanno ampliando le modalità con cui l'IA può supportare la ricerca scientifica e medica e la scoperta di farmaci. Gli LMM possono essere utilizzati in diversi ambiti della ricerca scientifica: pos-

[1] Etica e governance dell'intelligenza artificiale nello sviluppo dei farmaci. Ginevra: Organizzazione Mondiale della Sanità, in pubblicazione

sono generare il testo per un articolo scientifico, per la sottomissione di un lavoro o per la stesura di una peer-review [34]; possono essere utilizzati per riassumere testi, compresi i riassunti per gli articoli accademici, o per generare abstract.
Gli LMM possono anche essere usati per analizzare e riassumere i dati per ottenere nuove conoscenze nella ricerca clinica e scientifica. Possono essere usati anche per modificare un testo, migliorandone la grammatica, la leggibilità e la concisione per documenti scritti come articoli o richieste di sovvenzione. Inoltre, gli LMM possono essere utilizzati per ottenere approfondimenti partendo dai dati con cui sono stati addestrati [34]. Si sostiene che un LMM addestrato su milioni di articoli accademici sia in grado di analizzare la ricerca scientifica per rispondere a domande, estrarre informazioni o generare testi rilevanti [64].
Gli LMM sono utilizzati anche nella scoperta di farmaci e, in particolare, nella progettazione ex novo di farmaci per sviluppare nuove molecole con proprietà specifiche [65].

Rischi e sfide

Le principali riviste mediche e scientifiche hanno già reagito all'emergere dell'uso degli LMM, al loro potenziale e al loro impatto sulla ricerca scientifica. Per esempio, una casa editrice accademica ha stabilito due regole:

1. **un LMM non è accettato come autore accreditato in un articolo di ricerca**
2. **i ricercatori che utilizzano un LMM devono documentarne l'uso nella sezione sui metodi e nei ringraziamenti [66].**

La World Association of Medical Editors ha limitato la paternità di un articolo ai soli esseri umani [67]. Le preoccupazioni generali sull'uso degli LMM nella ricerca scientifica sono:

- **Mancanza di responsabilità:** la paternità di un articolo di ricerca scientifica o medica richiede una responsabilità che non può essere assunta da uno strumento di IA [66]. La mancanza di responsabilità è stata alla base della decisione di un importante editore accademico e della World Association of Medical Editors di non accettare un LMM come autore accreditato.
- **Pregiudizio a favore dei Paesi ad alto reddito:** la maggior parte della ricerca scientifica e medica utilizzata per istruire gli LMM è condotta nei Paesi ad alto reddito. Pertanto, è probabile che i risultati di una ricerca sugli LMM siano orientati verso la prospettiva dei Paesi ad alto reddito [34]. Ciò può rafforzare ed esacerbare la tendenza a ignorare e non citare le ricerche provenienti dai Paesi a basso o medio reddito [68], soprattutto se la pubblicazione è scritta in caratteri non latini.
- **Allucinazioni e/o disinformazione:** un LMM può avere allucinazioni riassumendo o citando articoli accademici o altre informazioni che non esistono [69].
- **Indebolimento della fiducia:** l'uso di un LMM per attività come la peer-review potrebbe minare la fiducia in tale processo [69].
- **Accessibilità degli LMM e delle conoscenze da essi generate:** come altri strumenti, tecnologie e informazioni utilizzati nella ricerca scientifica e medica, si prevede che gli LMM o alcune loro

funzioni siano disponibili a pagamento, con un costo che aggrava il divario digitale e di conoscenza e colpisce i ricercatori con minor sostegno e pochi finanziamenti che cercano di essere parte della ricerca medica e scientifica [34].

Sebbene l'IA (compresi gli LMM) possa favorire lo sviluppo di farmaci, vi è anche la preoccupazione per il suo utilizzo in questo settore, che viene esaminato in un'altra pubblicazione dell'OMS (vedi **nota 1** a pagina 38).

3. Rischi per i sistemi sanitari e la società e preoccupazioni etiche sull'uso degli LMM

I rischi e le preoccupazioni per l'utilizzo dei modelli linguistici di grandi dimensioni riguardano da una parte i singoli utenti direttamente coinvolti, come i professionisti sanitari, i pazienti, i ricercatori o i caregiver, e dall'altra parte l'intera collettività.
In particolare, i rischi emergenti associati all'implementazione degli LMM e di altre tecnologie basate sull'IA nell'assistenza sanitaria includono:

1. **rischi per il sistema sanitario di un Paese**
2. **rischi per la regolamentazione e la governance**
3. **rischi per l'impatto sociale.**

3.1 Sistemi sanitari

I sistemi sanitari si basano su sei elementi costitutivi: l'erogazione dei servizi, il personale sanitario, i sistemi di informazione sulla salute, l'accesso alle cure essenziali, i finanziamenti, la leadership e governance [70]. Gli LMM potrebbero avere un impatto diretto o indiretto su ciascuno di questi elementi. Di seguito vengono descritti i rischi associati all'uso degli LMM che potrebbero mettere alla prova, su più livelli, un sistema sanitario.

Sovrastimare i benefici degli LMM e ignorarne i rischi

La tendenza a sovrastimare e a sopravvalutare ciò che l'IA è in grado di fare può portare all'adozione di prodotti e di servizi che non sono stati sottoposti a una rigorosa valutazione di efficacia e sicurezza [1]. Ciò è dovuto in parte al forte fascino del "tecno soluzionismo", teoria che identifica nelle tecnologie come l'IA e gli LMM dei "proiettili magici" per eliminare le più profonde barriere sociali, strutturali, economiche e istituzionali [1], ancora prima che ne siano concretamente dimostrate utilità, sicurezza ed efficacia.
Gli LMM sono modelli nuovi e non testati. Non generano fatti ma informazioni che sembrano fatti e che possono però rivelarsi inaccurate. Questi modelli sono stati oggetto di un forte interesse da parte dei consumatori, della politica e dell'opinione pubblica. Tuttavia, potrebbero indurre i rappresentanti politici, gli erogatori dei servizi e i pazienti a sopravvalutare i loro benefici e a ignorare le sfide e i problemi che gli LMM possono introdurre. I politici potrebbero non ottenere i dati per determinare quanto ampio debba essere l'uso degli LMM prima che questi vengano effettivamente sviluppati e utilizzati.
L'impiego di un LMM non dovrebbe essere privilegiato rispetto all'uso di tecnologie basate sull'IA o di soluzioni analogiche già utilizzate, che possono non avere finanziamenti ma di cui è dimostrata l'efficacia terapeutica o l'utilità per la sanità pubblica.

Una politica sanitaria sbilanciata e investimenti non appropriati, soprattutto nei Paesi a basso e a medio reddito, possono distogliere l'attenzione e le risorse da interventi di comprovata efficacia, aumentando invece la pressione sui ministeri della salute per ridurre la spesa pubblica destinata alla salute [1].

Accessibilità e convenienza

Diversi fattori possono compromettere l'accesso equo agli LMM. Uno di questi è il divario digitale, che limita l'uso degli strumenti digitali a determinati Paesi, regioni o segmenti di popolazione. Il divario digitale porta alla creazione di ulteriori disparità, molte delle quali influenzano l'impiego dell'IA. Inoltre, l'IA stessa può a sua volta rafforzare ed esacerbare queste disparità. Un altro fattore che può compromettere l'accesso agli LMM è in molti casi l'obbligo di pagamento di un canone o di un abbonamento, a differenza di quanto avviene con Internet, perché lo sviluppo e la gestione di un LMM sono costosi. È stato infatti stimato che il funzionamento di ChatGPT richieda 700.000 dollari al giorno [71]. Alcune aziende stanno introducendo tariffe di abbonamento per le nuove versioni degli LMM [72]; questo potrebbe rendere alcuni LMM non accessibili, non solo per i Paesi a basso e a medio reddito, ma anche per i cittadini, i sistemi sanitari o i governi in contesti poveri di risorse all'interno dei Paesi ad alto reddito [54]. All'opposto, le persone meno abbienti, in tutti i Paesi, potrebbero ritrovarsi a utilizzare gli LMM come soluzione economicamente vantaggiosa per la salute, mentre l'accesso ai professionisti sanitari potrebbe essere riservato alle persone con maggiori disponibilità economiche. Un terzo fattore consiste nel fatto che al momento la maggior parte degli LMM opera solo in lingua inglese. Pertanto, anche se questi modelli possono ricevere input e fornire output in altre lingue, è più probabile che in queste siano generate informazioni false o fuorvianti [73].

Bias a livello di sistema

Come detto, l'insieme di dati utilizzati per addestrare i modelli di IA ha un bias di selezione, in quanto alcuni gruppi di persone sono meno rappresentati, come il genere femminile, le minoranze etniche, gli anziani, le comunità rurali e i gruppi di popolazione più svantaggiati.

In generale, l'IA è orientata verso le popolazioni per le quali si dispone di un maggior numero di dati; per questo motivo nelle società caratterizzate da maggiori disuguaglianze l'IA può svantaggiare i gruppi minoritari [1].

È probabile che i bias aumentino con la scala del modello, che può essere un problema soprattutto con gli LMM [74] perché la quantità di dati utilizzati per addestrare i modelli continua ad ampliarsi, anche se vengono sviluppati i cosiddetti piccoli modelli linguistici. Questi bias potrebbero introdurre discriminazioni in tutto il sistema sanitario, influenzando l'accesso delle persone ai servizi di base, o a quelli specialistici [75]. Allo stesso tempo, gli LMM sono portati a includere dati che portano a respingere forme di bias e stereotipi. I ricercatori hanno visto come richiedere a un modello di LMM di non affidarsi agli stereotipi abbia un notevole effetto positivo sulla risposta che viene fornita dall'algoritmo [74].

Impatto sul lavoro e sull'occupazione

Una banca d'investimento ha stimato che gli LMM comporteranno la perdita (o il "degrado") di almeno 300 milioni di posti di lavoro [76]. Un rapporto dell'Organizzazione per la cooperazione economi-

ca e lo sviluppo (OECD) ha analizzato come, all'interno dei Paesi membri, le professioni più a rischio di automazione guidata dall'IA siano quelle altamente qualificate, in parte proprio a causa dell'uso degli LMM, per cui "le professioni nel campo della finanza, della medicina e del diritto (…) possano improvvisamente trovarsi a rischio di automazione da parte dell'IA" [77]. Per molti Paesi, tuttavia, l'assistenza sanitaria non è un'industria, ma una funzione fondamentale del governo, per cui non è detto che i professionisti sanitari vengano sostituiti dalla tecnologia. Inoltre, molti Paesi continuano a lamentare forti carenze di professionisti sanitari [1], anche in seguito alla pandemia di COVID-19 [78]. L'OMS ha stimato che entro il 2030 vi sarà una carenza di 10 milioni di professionisti sanitari [79], soprattutto nei Paesi a basso e a medio reddito. Pertanto, gli LMM che venissero riconosciuti come sicuri ed efficaci potrebbero essere sfruttati proprio per colmare il divario tra la domanda e l'offerta della forza lavoro necessaria a fornire un'adeguata assistenza sanitaria.

Un'altra preoccupazione è l'impatto dell'introduzione degli LMM sul numero dei professionisti sanitari attuali e futuri. Un'importante azienda tecnologica ha stimato che fino all'80% dei posti di lavoro subirà conseguenze in seguito all'introduzione dell'IA [80]. La società di consulenza Accenture ha realizzato una stima sul fatto che il 40% delle ore lavorative potrà essere influenzato dagli LMM e ha osservato ottimisticamente che "l'impatto positivo sulla creatività e sulla produttività umana sarà consistente" [81]. Tuttavia, come già osservato, l'introduzione degli LMM potrebbe portare a sfide significative per molti professionisti sanitari, che dovranno essere formati e adattarsi agli LMM. I sistemi sanitari dovranno tenere conto delle sfide che ciò comporta per chi deve erogare l'assistenza e dei rischi per i pazienti e i caregiver.

Una terza preoccupazione riguarda il carico mentale e psicologico che ricade sulle persone responsabili della revisione dei contenuti degli LLM, dell'annotazione dei dati utilizzati per l'addestramento degli LMM e della rimozione di contenuti abusivi, violenti o mentalmente disturbanti. Coloro che si occupano di filtrare questi contenuti spesso risiedono in Paesi a basso e a medio reddito e percepiscono salari bassi, per cui possono andare incontro a stress e disturbi psicologici per la revisione dei dati, senza avere accesso a un supporto psicologico o ad altre forme di assistenza medica [73].

Dipendenza dei sistemi sanitari da LMM non adatti

Se da un lato gli LMM potrebbero contrastare la persistente carenza di professionisti sanitari e rendere i servizi più capillari, dall'altro questi ultimi potrebbero diventare dipendenti dagli LMM e in particolare dalle tecnologie LMM sviluppate dall'industria. Pertanto, se gli LMM utilizzati nell'assistenza sanitaria non venissero mantenuti e aggiornati, se venissero ridotti oppure se fossero aggiornati solo per l'uso in contesti ad alto reddito, i sistemi sanitari basati su questi modelli dovrebbero adattarsi e fornire un'assistenza sanitaria senza LMM. Ciò sarebbe molto difficile nel momento in cui i professionisti sanitari fossero stati dequalificati nella loro mansione e avessero delegato alcune responsabilità all'IA, oppure se i pazienti avessero fatto leva sul loro esclusivo utilizzo.
Un rischio correlato è che, nel caso in cui gli LMM non riuscissero a tutelare la privacy e la riservatezza del paziente, l'eccessiva dipendenza dagli LMM potrebbe minare la fiducia dei singoli e della collettività nel sistema sanitario, in quanto i cittadini non si sentirebbero più sicuri nell'accedere ai servizi sanitari senza mettere a rischio la propria privacy.

Rischi per la sicurezza informatica

Man mano che i sistemi sanitari diventano sempre più dipendenti dall'IA, le tecnologie potrebbero essere oggetto di cyberattacchi e di hacking. Alcuni servizi informatici del sistema sanitario potrebbero "andare giù" e non essere utilizzabili, con anche il rischio di una possibile manipolazione dei dati utilizzati per addestrare l'algoritmo e la conseguente modifica delle prestazioni e raccomandazioni. I dati inoltre potrebbero essere soggetti a furto per ottenere in cambio un riscatto economico [1]. Un rischio particolare per la sicurezza, già sottolineato in precedenza, riguarda l'alimentazione degli LMM con dati sensibili non destinati alla divulgazione o al loro uso non autorizzato.
In aggiunta, gli stessi potrebbero a loro volta essere soggetti ad attacchi hacker con rischi per la sicurezza informatica, come per esempio la "prompt injection", un attacco in cui i dati vengono immessi in un LMM da una terza parte, per indurre un comportamento del modello non voluto dallo sviluppatore [82].
Una prompt injection potrebbe, per esempio, istruire un LMM progettato per rispondere a domande su un database a cancellare informazioni dal database o a modificarle. Non c'è ancora una soluzione per risolvere questa falla. Anche se le prompt injection vengono attualmente impiegate dagli esperti della sicurezza informatica come esempio per illustrare le sfide degli LMM, questi attacchi potrebbero essere concretamente utilizzati da malintenzionati per rubare dati o realizzare tentativi di frode nei confronti degli utenti [83].

3.2 Conformità ai requisiti regolatori e legali

Sebbene in futuro possano essere emanate nuove leggi per regolamentare l'uso dell'IA, alcuni atti legislativi e regolamenti già esistenti, come in particolare le leggi sulla protezione dei dati e gli obblighi internazionali in materia di diritti umani, sono applicabili allo sviluppo e alla diffusione degli LMM.
Alcuni LMM, così come attualmente realizzati e introdotti per l'uso pubblico, possono violare alcune leggi fondamentali sulla protezione dei dati, come il *Regolamento generale sulla protezione dei dati dell'Unione Europea* (GDPR) [84], che riconosce vari diritti, come la protezione rispetto ai processi decisionali automatizzati. Risulta invece evidente come tali diritti, tutele e requisiti debbano guidare lo sviluppo dell'IA [85].
Alcune di queste violazioni hanno portato a indagini sugli LMM all'interno degli Stati membri dell'Unione Europea [83] e al di fuori dei suoi confini, come in Canada [86]. Le violazioni hanno incluso:

1. **LMM che hanno raccolto da Internet e utilizzato i dati personali senza il consenso delle persone (e senza un "interesse legittimo" per la raccolta di tali dati)** [87]
2. **LMM che non erano in grado di informare le persone che stavano utilizzando i loro dati o che non riconoscevano il diritto dell'utente di correggere eventuali errori, di cancellare i propri dati (il "diritto all'oblio") o di opporsi all'uso di tali dati** [87]
3. **LMM che non erano completamente trasparenti nell'uso dei dati sensibili forniti a un chatbot o ad altre interfacce per i consumatori, sebbene, per legge, un utente debba essere in grado di cancellare i dati immessi nella chat** [83]
4. **LMM che non disponevano di un adeguato sistema di age-gating per filtrare gli utenti di età**

inferiore ai 13 anni e quelli di età compresa tra i 13 e i 18 anni per i quali non era stato fornito il consenso dei genitori [88]

5. **LMM che non erano in grado di prevenire la violazione dei dati personali** [87]
6. **LMM che pubblicavano informazioni personali inesatte, dovute in parte ad "allucinazioni"** [89].

Altre possibili violazioni del Regolamento generale sulla protezione dei dati includono il requisito del "diritto alla spiegazione", in base al quale un ente che utilizza i dati personali per il trattamento automatizzato deve spiegare come il sistema, in questo caso un LMM, prenda le decisioni.[1]
Come già detto, le aziende tecnologiche non sono ancora in grado di spiegare come gli LMM prendano le decisioni, anche se alcune stanno lavorando su approcci nuovi per soddisfare il requisito della "spiegabilità" [90].
Molte violazioni sono di notevole importanza, in particolare quelle che riguardano le modalità di allenamento degli LMM, il loro utilizzo e la loro gestione da parte dei data controller.

È possibile che gli LMM non diventino mai conformi al *Regolamento generale sulla protezione dei dati* o ad altre leggi sulla protezione dei dati [91].
Un reclamo presentato nel 2023 all'Autorità per la protezione dei dati di uno Stato membro dell'Unione Europea sosteneva che l'LLM di un'azienda tecnologica e l'approccio per il quale è stato sviluppato e viene gestito violi sistematicamente il *Regolamento generale sulla protezione dei dati* [92].
Inoltre, molte violazioni della protezione dei dati possono anche violare le leggi sulla protezione dei consumatori [93]. Più in generale, se questi problemi non possono essere risolti, si contravviene anche ai principi guida dell'OMS sull'uso dell'IA nel campo della salute, compresi i principi di protezione dell'autonomia e di garanzia di trasparenza, "spiegabilità" e comprensibilità.

L'incapacità delle aziende tecnologiche di rispettare le leggi esistenti è il motivo per cui alcune hanno espresso una seria preoccupazione per le imminenti normative sull'IA. In risposta alla prevista introduzione da parte dell'Unione Europea di una legge sull'IA, l'amministratore di una grande azienda tecnologica ha infatti dichiarato che potrebbe non offrire il suo prodotto chiave LMM in Europa, perché non in grado di rispettare le normative [94]. Questo ultimatum potrebbe portare a una riduzione dei diritti alla privacy e di altre tutele per la salute, che dipenderanno dalla volontà di un Paese di rinunciare ad alcuni diritti umani.

3.3 Rischi e preoccupazioni per la società

Come per altre tecnologie nel campo dell'IA, si prevede che gli LMM abbiano impatti sulla società che vanno oltre al sistema sanitario e che non possono essere governati da una singola legge o policy. Tra questi, la possibilità che gli LMM rafforzino il potere e l'autorità di un esiguo gruppo di aziende tecnologiche (e dei loro dirigenti) che si trovano in prima linea nella commercializzazione degli LMM.
Inoltre, gli LMM possono avere un impatto negativo a livello ambientale e climatico, per l'elevato consumo di elementi, come carbonio e acqua, richiesti per il continuo miglioramento e l'implemen-

[1] Vedi Articolo 15(1)(h) e Comma 71 del GDPR

tazione delle tecnologie stesse. Tali tecnologie diventano rapidamente "presenti nella vita di miliardi di persone a un ritmo più rapido della capacità delle culture di assorbirle in modo sicuro" [4], compresi i campi sanitario e medico, prima ancora che gli esseri umani possano garantire che le tecnologie dell'IA non rimpiazzino la nostra autorità epistemica con informazioni, prove e raccomandazioni che sono spesso non accurate, false, gravate da bias e sprovviste di morale o di un inquadramento contestuale. Un'altra grande preoccupazione è che gli LLM aumentino la violenza di genere facilitata dalla tecnologia, compresi il cyber-bullismo, l'incitamento all'odio, l'uso di immagini e video senza consenso, come il deep fake. Quest'ultimo non viene affrontato in questo report, ma merita maggiore considerazione da parte dell'OMS, per le implicazioni negative molto gravi per la salute e il benessere di popolazioni, come le ragazze e le donne, che sono bersaglio di tali utilizzi dell'IA.

Sfide per le grandi aziende tecnologiche

L'emergere degli LMM, che continuano ad accrescersi con un numero sempre maggiore di parametri, ha rafforzato il dominio e la centralità di poche grandi aziende tecnologiche che sviluppano e implementano l'IA [96]. Poche aziende e governi hanno le risorse umane e finanziarie, le competenze, i dati e la potenza di calcolo per sviluppare LMM sempre più sofisticati [96]. La potenza di calcolo e gli investimenti negli LMM sono aumentati e, con la crescita della domanda di IA, il reclutamento di "talenti dell'IA" è costoso [97, 98]. Gli LMM che contengono i microchip più potenti disponibili richiedono molti computer e migliaia di chip che lavorano insieme per funzionare, con i computer che lavorano senza sosta per settimane o addirittura mesi [99].

Con il continuo aumento dei costi di formazione, impiego e manutenzione degli LMM, c'è il rischio di un dominio industriale da parte di poche aziende di quello che è potenzialmente un elemento costitutivo di molti prodotti e servizi (anche nel settore sanitario), a tal punto da escludere le università, le start-up e persino i governi [100]. Nella ricerca sull'IA ci sono già prove convincenti del fatto che le aziende più grandi stiano escludendo sia le università sia i governi. Un'indicazione è data dal luogo di lavoro scelto dai diplomati con un dottorato in IA. Le percentuali di coloro che scelgono di lavorare in azienda sono senza precedenti. Mentre nel 2004 solo il 20% circa entrava nell'industria, nel 2020 questo numero è arrivato a più del triplo, circa il 70% [101]. Professori e altri accademici specializzati nell'IA vengono portati via dalle università per lavorare nell'industria, dove il loro numero è aumentato di otto volte dal 2006, non solo negli Stati Uniti ma anche in altri Paesi [101]. L'industria è anche arrivata a dominare sia i governi sia il mondo accademico per quanto riguarda la potenza di calcolo e l'uso di grandi dataset. Nel 2021, i modelli dell'industria erano 29 volte più grandi di quelli accademici [101]. Inoltre, gli investimenti globali, soprattutto da parte dei governi dei Paesi ad alto reddito sono molto inferiori a quelli dell'industria. Come si legge in uno studio: "Nel 2021, le agenzie governative statunitensi non legate alla difesa hanno stanziato 1,5 miliardi di dollari per l'IA. Nello stesso anno, la Commissione Europea prevedeva di spendere 1 miliardo di euro. A livello globale, invece, l'industria ha investito più di 340 miliardi di dollari per l'IA nel 2021, superando di gran lunga gli investimenti pubblici" [101].

Se le grandi aziende tecnologiche possiedono il dominio degli "input dell'IA", significa che dominano anche gli output e i risultati. La quota industriale dei maggiori modelli di IA è passata dall'11% nel 2010 al 96% nel 2021, mentre il numero di report di ricerca con uno o più coautori dell'industria è

aumentato del 16% tra il 2000 e il 2020 [101]. Il dominio delle grandi aziende tecnologiche definisce non solo le applicazioni e gli usi dell'IA, ma anche, in misura crescente, le priorità della ricerca iniziale [101]. Con il dominio dell'industria e la mancanza di investimenti governativi, si riducono le alternative per importanti tecnologie dell'IA di interesse pubblico, tra cui l'assistenza sanitaria e la medicina. A differenza del settore farmaceutico, per esempio, dove vi sono notevoli investimenti governativi, no-profit e filantropici nella ricerca e nello sviluppo, soprattutto nelle prime fasi critiche dello sviluppo di un farmaco e anche per lo sviluppo successivo di alcuni farmaci [102]. Le aziende tecnologiche avranno quindi una sorta di supervisione sul funzionamento dei sistemi che sono alla base delle nostre economie e dei settori sociali, compresa l'assistenza sanitaria, sollevando preoccupazioni sulla capacità dei cittadini e degli organi di controllo di gestire le loro vite [101].
In assenza di alternative e di una regolamentazione (che potrebbe richiedere diversi anni per essere pienamente attuata, anche se le leggi sono state rese esecutive nel 2023), il modo in cui le grandi aziende tecnologiche prendono decisioni interne e il modo in cui si relazionano con le società e i governi diventano sempre più rilevanti. Le aziende potrebbero iniziare ad affrontare le varie problematiche, in collaborazione, per esempio, con il Frontier Model Forum [103] o con i governi dei Paesi ad alto reddito, inclusi diversi impegni presi volontariamente con il governo degli Stati Uniti [104] e con l'Unione Europea [105].

Un'altra preoccupazione è che le aziende non mantengano un impegno aziendale in materia di etica. Per esempio, le principali aziende tecnologiche hanno messo in secondo piano o eliminato i loro team etici, istituiti per garantire che la progettazione e lo sviluppo di modelli di IA rispettassero i principi etici interni [106], creando attriti che avrebbero richiesto all'azienda di rallentare o interrompere alcune attività di sviluppo, considerate potenzialmente rischiose. L'eliminazione di interi team che si occupano di questioni etiche per l'IA significa che i principi etici non sono "strettamente legati alla progettazione del prodotto" [108], lasciando così un vuoto da colmare.

Diverse grandi aziende tecnologiche, attraverso il Frontier Model Forum, si sono impegnate a garantire "uno sviluppo responsabile e sicuro dei modelli di IA avanzati", compresi gli LMM, "identificando le migliori pratiche per lo sviluppo e l'impiego responsabile dei modelli di frontiera" e "collaborando con i responsabili politici, gli accademici, la società civile e le aziende per condividere le conoscenze riguardo alla fiducia e ai rischi per la sicurezza" [103]. Le aziende tecnologiche si sono impegnate volontariamente nei confronti del governo degli Stati Uniti a evitare pregiudizi e discriminazioni dannose e a proteggere la privacy [104]. Non è chiaro, tuttavia, se gli impegni volontari o i partenariati saranno sufficienti a sostituire un solido impegno in materia di etica. In un caso, per esempio, il team etico di un'azienda, che aveva raccomandato di interrompere il rilascio di un nuovo LMM, ha modificato i propri documenti, minimizzando i rischi precedentemente segnalati [106]. Le grandi aziende tecnologiche non hanno una storia né sono specializzate nello sviluppo di prodotti e servizi sanitari, pertanto potrebbero non essere sensibili alle esigenze dei sistemi sanitari, degli erogatori e dei pazienti e potrebbero non occuparsi, per esempio, di privacy o di garanzia della qualità, aspetti che sono propri alle aziende sanitarie tradizionali e ai fornitori di servizi sanitari pubblici. La loro sensibilità può migliorare con il tempo, come è avvenuto in altre aziende che hanno fornito beni e servizi sanitari per diversi decenni.

Molte aziende che sviluppano LMM non sono trasparenti né con i governi né con le autorità regolatorie, né con le aziende che potrebbero utilizzare i loro modelli e che possono richiedere prove, dati, prestazioni e altre informazioni per valutare i rischi e i benefici di un LMM [97,106] e il numero di parametri nel modello, che è un indicatore della sua potenza [8]. Le aziende che usano tali modelli per sviluppare i propri prodotti e servizi a loro volta non rivelano come valutano i rischi e gli aspetti etici, le tutele da porre in atto, la reazione degli LMM a tali tutele e quando l'uso della tecnologia dovrebbe essere limitato o interrotto. Il Foundation Model Transparency Index, che valuta in base a cento indicatori i dieci sviluppatori leader di LLM, mostra che "nessuno dei principali sviluppatori di modelli di base è vicino a fornire una trasparenza adeguata, rivelando una fondamentale mancanza di trasparenza nel settore dell'IA" [107].

L'accordo volontario tra il Governo federale degli Stati Uniti e alcune grandi aziende tecnologiche prevede due impegni di trasparenza. Le aziende si impegnano a:

1. **condividere le informazioni sui rischi di gestione con l'industria, i governi, la società civile e il mondo accademico**
2. **rendere pubbliche le capacità, i limiti e le aree di utilizzo appropriato e inappropriato dei loro sistemi di IA [104].**

Sebbene questi impegni possano essere considerati un miglioramento rispetto allo status quo, sono volontari e aperti all'interpretazione da parte di ciascuna azienda tecnologica, che potrebbe non essere adeguata sul piano della trasparenza in assenza di requisiti normativi concreti.
Le aziende tecnologiche si stanno affrettando a rendere disponibili i nuovi LMM il più rapidamente possibile per le pressioni commerciali interne o per la concorrenza [106], prima di comprendere appieno il funzionamento degli LMM [109] e indipendentemente dal fatto che siano stati identificati e gestiti i test, le tutele e i rischi etici [106,110]. Un dirigente di queste aziende ha dichiarato che è un "errore fatale in questo momento preoccuparsi di aspetti che potranno essere risolti in seguito" [106]. Le aziende tecnologiche cercano il vantaggio di essere le prime a rilasciare un certo LLM, perché la disponibilità di LLM in alcuni settori, come la ricerca in Internet, comporta guadagni. Secondo una di queste aziende tecnologiche, ogni 1% di quota di mercato in un motore di ricerca equivale a 2 miliardi di dollari di entrate aggiuntive [108]. Un dirigente di una grande azienda tecnologica ha osservato che il suo LMM "non è perfetto", ma che verrà rilasciato perché "il mercato lo richiede" [8]. Le aziende che rilasciano gli LMM senza aver pienamente identificato, valutato, preso in carico e mitigato i rischi accumulano un "debito etico", le cui conseguenze finali saranno subite non dalle aziende ma da coloro che sono più vulnerabili agli impatti negativi di tali tecnologie [109]. I membri del Frontier Model Forum si sono impegnati a "far progredire la ricerca sulla sicurezza dell'IA" e a "identificare le migliori pratiche" [103], e gli impegni volontari assunti nei confronti del Governo degli Stati Uniti includono test interni ed esterni dei sistemi di IA prima del loro rilascio [104].

La pressione commerciale può portare le aziende tecnologiche non solo ad affrettare l'immissione sul mercato degli LMM, ma anche a retrocedere come priorità o addirittura abbandonare prodotti e servizi che potrebbero avere un beneficio significativo per la salute pubblica, per privilegiare servizi più remunerativi. Nel 2023, una grande azienda tecnologica ha licenziato un team che aveva

sviluppato un LMM chiamato ESMFold, un Protein Language Model in grado di predire una struttura proteica a livello atomico completa a partire da una singola sequenza e che ha generato un database di oltre 600 milioni di strutture proteiche. Si teme che l'azienda non sia disposta ad "assorbire i costi per mantenere in funzione il database e un altro servizio che consente ai ricercatori di utilizzare questo algoritmo ESM (Evolutionary Scale Modeling) su nuove sequenze proteiche" [111].

L'impronta ecologica degli LMM su carbonio e acqua

Un'altra conseguenza della crescente dimensione degli LMM è il loro impatto ambientale. Gli LMM richiedono grandi quantità di dati e il training con i dati richiede quantità significative di energia [112]. In una grande azienda tecnologica l'addestramento di un nuovo LMM ha consumato circa 3,4 GWh in due mesi, pari al consumo annuale di energia di 300 famiglie statunitensi [112]. Mentre alcuni LMM vengono addestrati in centri dati che utilizzano energia rinnovabile o a zero emissioni, la maggior parte dei modelli di IA viene addestrata con reti elettriche alimentate da combustibili fossili [112]. Il consumo di elettricità continuerà ad aumentare via via che più aziende produrranno nuovi LMM e ciò potrebbe incidere in modo significativo sul cambiamento climatico.

L'OMS considera il cambiamento climatico una sfida urgente per la salute globale, che richiede un'azione prioritaria, da ora e nei decenni a venire. Tra il 2030 e il 2050 si prevede che il cambiamento climatico causerà circa 250.000 morti in più all'anno nel mondo a causa di malnutrizione, malaria, diarrea e aumento della temperatura. Il costo dei danni diretti alla salute entro il 2030 è stimato in 2-4 miliardi di dollari all'anno. Le aree con servizi sanitari meno efficienti, la maggior parte delle quali si trova nei Paesi a basso e a medio reddito, saranno quelle con minore capacità di affrontare la situazione senza l'assistenza necessaria per prepararsi e rispondere [1].

Gli LMM hanno anche un notevole impatto idrico. L'addestramento di un primo LMM in una grande azienda tecnologica ha consumato 700.000 litri di acqua, e sarebbero stati anche di più in altri centri di elaborazione dati [113]. Sebbene molti sviluppatori siano sempre più consapevoli della loro impronta di carbonio, molti non sono consapevoli della loro impronta idrica [114]. Una breve conversazione di un LMM (tra 20 e 50 domande e risposte) richiede l'equivalente di una bottiglia d'acqua da mezzo litro. L'impatto idrico complessivo dell'addestramento di un LMM, che comprende tutta l'acqua consumata, compresa la produzione di server IA, il trasporto e la fabbricazione di chip, può essere anche significativamente maggiore [114].

I centri di elaborazione dati possono mettere a dura prova le riserve idriche locali. Per esempio, il centro dati di un'azienda tecnologica ha utilizzato più del 25% di tutta l'acqua consumata da una città dell'Oregon, negli Stati Uniti [114]. Un'altra grande azienda tecnologica sta pianificando la costruzione di un centro dati in un Paese che sta attraversando una grave siccità, tanto che i residenti sono costretti a usare acqua salata per il consumo umano [115]. Tenere traccia dell'impatto idrico è difficile perché, mentre c'è una maggiore consapevolezza, misurazione e trasparenza sull'impronta carbonica, le aziende non sono altrettanto trasparenti sulla loro impronta idrica o non la misurano [114].

Algoritmi "pericolosi" che sostituiscono l'autorità epistemica degli esseri umani

Un rischio sociale più generale associato all'emergere degli LMM è che, fornendo risposte plausibili che sono sempre più considerate una fonte di conoscenza, gli LMM possano alla fine minare l'autorità epistemica degli esseri umani, anche nei settori della sanità, della scienza e della medicina. Gli LMM, infatti, non generano conoscenza, non capiscono ciò che "dicono" e non fanno alcun ragionamento morale o contestuale nel rispondere alle domande.

Se le preoccupazioni sono reali, le società potrebbero non essere preparate alle conseguenze dei ragionamenti generati dai computer. Le prime forme di IA che fornivano informazioni attraverso gli algoritmi dei social media hanno diffuso disinformazione, con impatti negativi sulla salute mentale e un aumento della polarizzazione e delle divisioni [4]. Anche se le aziende tecnologiche lanciano ripetuti avvertimenti sui pericoli degli LMM, continuano a rilasciarli direttamente al pubblico, senza tutele o controlli normativi, in modi che potrebbero non solo sostituire il controllo umano sulla produzione di conoscenza, ma anche ridurre la capacità degli esseri umani di usare la conoscenza in modo sicuro nell'assistenza sanitaria, nella medicina e in altri ambiti da cui le società dipendono. Tali pericoli potrebbero riguardare in particolare le persone e le comunità in contesti poveri di risorse, poiché è improbabile che i loro dati siano stati utilizzati per addestrare un sistema di IA, riducendo così l'accuratezza delle risposte. D'altra parte tali gruppi potrebbero seguire i consigli di un sistema di IA, in particolar modo se non è disponibile un professionista della salute o un medico per contestualizzare o correggere una risposta falsa o imprecisa generata da un LMM.

L'immissione nel pubblico dominio di informazioni sempre più imperfette o la disinformazione da parte degli LMM potrebbe alla fine portare a un "collasso del modello", in cui gli LMM addestrati su informazioni inaccurate o false inquinano anche le fonti pubbliche di informazione, come Internet [116, 117]. Per evitare questo scenario e massimizzare i benefici degli LMM nell'assistenza sanitaria e in altre aree di importanza sociale, i governi, la società civile e il settore privato devono orientare queste tecnologie verso il bene comune.

→ Torna all’indice

II. Etica e governance degli LMM nella sanità e nella medicina

I principi etici definiti dal gruppo di esperti dell'OMS (vedi prima) forniscono un orientamento ai portatori di interesse per i requisiti etici di base che dovrebbero guidare le loro decisioni e azioni nello sviluppo, implementazione e valutazione dell'uso degli LMM nella sanità e nella medicina. Per i governi, le agenzie del settore pubblico, i ricercatori, le aziende e gli implementatori questi principi dovrebbero essere la base per come gestire l'uso degli LMM.

La governance comprende le funzioni di indirizzo e di regolamentazione da parte dei governi e di altri decisori, comprese le agenzie internazionali per la salute, per porre in atto una politica sanitaria nazionale e globale che favorisca la copertura sanitaria universale. La governance è anche un processo politico che comporta il bilanciamento di influenze e richieste contrastanti [1]. Le leggi e le politiche attuali sono insufficienti per una gestione efficace dell'uso degli LMM, poiché molte sono state scritte quando ancora non erano stati pubblicati i primi LMM. La governance degli LMM, come la governance generale dell'IA, comporta l'applicazione di leggi e regolamenti attuali e nuovi, di una norma sui principi etici, di obblighi di rispetto dei diritti umani, di codici per la pratica e le procedure interne delle aziende tecnologiche, delle associazioni di settore e degli enti che devono definire gli standard.

Attualmente, gli LMM vengono implementati più rapidamente rispetto alla nostra capacità di comprenderne appieno le capacità e le fragilità. All'inizio era stato anche proposto per affrontare le preoccupazioni riguardo agli LMM di porre un divieto o una moratoria sul loro sviluppo [118]. Sebbene alcuni Paesi limitino l'uso o addirittura vietino alcuni LMM, la maggior parte dei governi ora cerca di garantire che il loro utilizzo sia orientato verso risultati socialmente utili grazie a una governance appropriata. Anche le principali aziende di IA hanno chiesto uno sviluppo attento, deliberativo degli LMM e di altre forme di IA. Né i governi né le aziende, tuttavia, sono immuni dalla competizione. Diversi governi sono impegnati in una "corsa agli armamenti" per la supremazia tecnologica, mentre anche le aziende di IA che chiedono regolamentazione sono guidate dalla pressione commerciale [119]. Mentre gli ottimisti ritengono che molte delle sfide e dei rischi dell'IA possano essere affrontati attraverso la progettazione, incluso l'utilizzo di set di dati sempre più ampi e algoritmi più potenti, i critici hanno sottolineato che i limiti degli LMM sono sistemici e che l'aumento delle dimensioni dei dati di addestramento e dei parametri dei modelli non supererà le lacune ma di fatto le amplificherà [59].

La governance degli LMM deve tenere il passo con il rapido sviluppo e il crescente utilizzo e non dovrebbe privilegiare né i governi che cercano un vantaggio tecnologico né le aziende che cercano un guadagno commerciale. I suggerimenti e le raccomandazioni che seguono, pongono i principi etici

e gli obblighi in materia di diritti umani al centro di una governance appropriata, comprendendo sia le procedure e le pratiche che dovrebbero essere introdotte dalle aziende, sia le leggi e le politiche poste in atto dai governi.

Gli LMM possono essere considerati prodotti di una serie (o catena) di decisioni riguardo alla loro programmazione e sviluppo da parte di uno o più attori. Le decisioni prese in ciascuna fase della catena del valore dell'IA possono avere conseguenze dirette e indirette su coloro che partecipano allo sviluppo, alla distribuzione e all'uso degli LMM. Le decisioni possono essere influenzate e regolate dai governi che promulgano e applicano leggi e politiche a livello nazionale, regionale e globale. La catena del valore dell'IA inizia con l'integrazione di diversi input, che comprendono l'infrastruttura dell'IA, come i dati, la potenza di calcolo e le competenze dell'IA, fino allo sviluppo di modelli di base con finalità generali. Questi modelli possono essere utilizzati direttamente da un utente per svolgere varie attività, spesso compiti non previsti per un dato modello (compresi quelli relativi all'assistenza sanitaria). Alcuni modelli di base per finalità generali sono addestrati specificamente per l'uso nell'assistenza sanitaria e nella medicina.

La governance appropriata degli LMM utilizzati nell'assistenza sanitaria e nella medicina dovrebbe essere definita in ciascuna fase della catena del valore, dalla raccolta dei dati all'implementazione delle applicazioni nell'assistenza sanitaria. Pertanto, le tre fasi critiche della catena del valore dell'IA discusse sono:

- **la progettazione e lo sviluppo di modelli di base con finalità generali (fase di progettazione e sviluppo)**
- **la definizione di un servizio, un'applicazione o un prodotto con un modello di base per finalità generali (fase di fornitura)**
- **l'implementazione di un'applicazione di servizio per l'assistenza sanitaria o di un servizio (fase di implementazione).**

In ciascuna fase della catena del valore dell'IA, bisogna porsi le seguenti domande.

- **Quale attore (il programmatore, il fornitore e/o l'implementatore) è nella posizione migliore per gestire i rischi rilevanti? Quali rischi dovrebbero essere gestiti nella catena del valore dell'IA?**
- **Come possono l'attore/gli attori principali affrontare tali rischi? Quali principi etici devono rispettare?**
- **Qual è il ruolo del governo nella gestione dei rischi? Quali leggi, politiche o investimenti potrebbe introdurre o applicare un governo per chiedere agli attori della catena del valore dell'IA di rispettare i principi etici?**

Durante la fase di progettazione e sviluppo, l'attenzione è focalizzata sulle pratiche che gli sviluppatori possono introdurre per soddisfare gli impegni etici, rispettare le norme e le politiche governative e gli investimenti. Durante la fase di fornitura, il focus è sulle misure che i governi possono introdurre per valutare e regolare l'uso degli LMM nell'assistenza sanitaria e nella medicina. Durante la fase di implementazione, infine, occorre utilizzare misure da parte dei governi e di tutti gli attori della catena del valore per garantire che i danni potenziali o effettivi agli utenti siano identificati ed evitati.

4. Progettazione e sviluppo di modelli di base per finalità generali

I modelli di base per finalità generali sono di solito addestrati su una grande quantità di dati, richiedendo una enorme potenza di calcolo. Lo sviluppo degli LMM richiede anche risorse umane specializzate, incluse competenze scientifiche e ingegneristiche. La guida dell'OMS sull'etica e la governance dell'IA per la salute [1] raccomanda che gli sviluppatori di IA medica "investano in misure per migliorare la progettazione, il controllo, l'affidabilità e l'autoregolamentazione dei loro prodotti". Anche se la maggior parte dei dati e delle raccomandazioni riportate di seguito potrebbe applicarsi a tutti i modelli di base per finalità generali, la guida è destinata ai modelli che possono essere o sono utilizzati nella sanità e nella medicina (direttamente da un utente o tramite un'applicazione o un servizio). Le raccomandazioni riportate di seguito sono anche destinate a guidare la progettazione e l'uso degli LMM addestrati specificamente per l'uso nella sanità e nella medicina, che possono essere utilizzati direttamente dagli utenti o attraverso un'applicazione o un servizio.

4.1 Rischi che devono essere gestiti durante lo sviluppo di modelli di base per finalità generali

La progettazione e lo sviluppo di modelli di base con finalità generali possono introdurre gravi rischi che, se non gestiti, potrebbero avere un impatto generalizzato sulla società o conseguenze negative specifiche sugli utenti di un LMM. L'eliminazione o la mitigazione dei rischi è responsabilità dello sviluppatore e – poiché è solo lo sviluppatore che può prendere certe decisioni durante la progettazione e lo sviluppo – tali rischi sono al di là del controllo dei fornitori e dei distributori che possono utilizzare l'algoritmo [120]. Le decisioni si riferiscono, per esempio, ai dati utilizzati per addestrare un LMM [121]. Gli obblighi di garantire la protezione e la qualità dei dati e di mitigare i bias sono al di fuori del controllo dello sviluppatore [121], così come le misure che dovrebbero essere introdotte per garantire che gli LMM non provochino una "tossicità alimentata dall'IA" [122]. La mancanza nel rendere gli sviluppatori degli LMM responsabili di tali difetti di progettazione protegge le aziende con più risorse, come ha osservato un rapporto, dalla "responsabilità di affrontare problemi (…) che i loro metodi potrebbero avere, nella fretta di dominare una nuova forma di IA applicata" [122]. Almeno otto rischi dovrebbero essere gestiti dallo sviluppatore di un modello di base per finalità generali, anche attraverso leggi e regolamenti governativi:

- **bias (associato alla progettazione e ai dati di addestramento)**
- **privacy (dei dati di addestramento e di altri dati di input)**
- **problemi di lavoro (filtraggio esternalizzato dei dati per rimuovere contenuti offensivi)**

- **impronta carbonica e idrica**
- **informazioni false, incitamento all'odio o disinformazione**
- **sicurezza e cybersicurezza**
- **preservazione dell'autorità epistemica degli esseri umani**
- **controllo esclusivo degli LMM.**

4.2 Misure che gli sviluppatori possono adottare per gestire i rischi con i modelli di base per finalità generali

Uno sviluppatore potrebbe adottare molte misure o pratiche per gestire tali rischi, sia come impegno nei confronti dei principi etici o delle leggi, sia per soddisfare i requisiti governativi.

Competenze di IA (personale scientifico e ingegneristico): uno sviluppatore può garantire che il suo personale scientifico e di programmazione sia in grado di identificare ed evitare i rischi. Le linee guida etiche dell'OMS [1] hanno formulato diverse raccomandazioni per la formazione del personale scientifico e ingegneristico e sull'inclusività del processo di progettazione. In particolare, il gruppo di esperti dell'OMS ha raccomandato agli sviluppatori di considerare i "requisiti di licenza o una certificazione per i programmatori di IA ad alto rischio, compresa l'IA per la salute".

Le aziende e altre entità che sviluppano un LMM che potrebbe essere usato nell'assistenza sanitaria, nella ricerca scientifica o nella medicina, dovrebbero considerare una certificazione o seguire una formazione per allinearsi ai requisiti per la professione medica e anche per aumentare la fiducia nei propri prodotti e servizi [1]. Eventuali norme, introdotte e applicate sia dagli sviluppatori sia dalle aziende, dovrebbero essere emanate dalle agenzie regolatorie e dovrebbero essere conformi al principio etico dell'OMS di promuovere il benessere umano, la sicurezza e l'interesse pubblico. Gli sviluppatori che non intendono, ma che prevedono che il loro LMM possa essere usato in ambito sanitario, dovrebbero garantire competenze interne per prevedere e affrontare tali utilizzi.

Dati: mentre risorse umane e potenza di calcolo sono essenziali per lo sviluppo degli LMM, i dati sono probabilmente il requisito infrastrutturale più critico. La qualità e il tipo dei dati utilizzati per addestrare un LMM determinano se esso rispetti i principi etici fondamentali e i requisiti legali [123]. Anche se, in sondaggi qualitativi, gli sviluppatori di IA concordano che la qualità dei dati conta e richiede un impegno significativo di tempo, il lavoro relativo ai dati è spesso sottovalutato e ciò può avere significative ripercussioni negative per l'IA in settori ad alto rischio come l'assistenza sanitaria e la medicina [123]. Se i dati non sono della qualità appropriata o se c'è un bias di selezione, vari principi guida dell'OMS potrebbero essere violati e tra questi la promozione del benessere umano, della sicurezza, dell'interesse pubblico e il principio di garantire l'inclusività e l'equità.

Siccome l'uso dei dati per l'assistenza sanitaria probabilmente richiederà una rigida aderenza alle leggi sul consenso informato, gli sviluppatori che addestrano LMM destinati all'uso nell'assistenza sanitaria e nella medicina potrebbero dover fare affidamento su insiemi di dati più piccoli [59]. Inoltre, il ricorso a insiemi di dati più piccoli potrebbe essere preferibile per garantire la qualità dei

dati e che i dati siano diversificati al fine di evitare bias [59] e che essi riflettano la composizione e la realtà delle popolazioni servite dall'LMM. Di contro, insiemi di dati più piccoli potrebbero aumentare il rischio di identificazione delle persone, che potrebbe esporle a un danno immediato o futuro. Fare affidamento su insiemi di dati più piccoli potrebbe avere come ulteriori benefici la riduzione delle impronte di carbonio e idrica [112] dei modelli e anche la possibilità per strutture più piccole di partecipare o sviluppare LMM che richiedono minori risorse dati, di calcolo, umane e finanziarie [59]. Indipendentemente dalle dimensioni dell'insieme di dati, gli sviluppatori prima di elaborare i dati devono effettuare "valutazioni dell'impatto sulla protezione dei dati", come richiesto dal *Regolamento generale sulla protezione dei dati*, considerando i rischi connessi al trattamento dei dati riguardo ai diritti e la libertà degli individui e il loro impatto sulla protezione dei dati personali [1]. La raccolta di dati da Paesi a basso e a medio reddito potrebbe configurare un "colonialismo dei dati", in cui i dati vengono utilizzati per scopi commerciali o non commerciali senza il dovuto rispetto del consenso, della privacy o dell'autonomia [1].

Le valutazioni potrebbero estendersi oltre ai rischi per la privacy e includere la qualità dei dati, per esempio se sono privi di bias e accurati. I ricercatori nel campo dell'IA che esaminano o verificano i set di dati sono riluttanti a investire il loro tempo per la verifica dei dati perché mentre la creazione di un set di dati per l'IA è semplice, la revisione è difficile, richiede tempo ed è costosa. Come ha dichiarato un ricercatore: "Fare il lavoro sporco è solo molto più difficile" [124].

Gli sviluppatori possono adottare altre misure per migliorare la qualità dei dati e rispettare le leggi sulla protezione dei dati. Indipendentemente dalle dimensioni del modello, dovrebbero, a differenza di come sono stati sviluppati i primi LMM, addestrare gli LMM su dati raccolti secondo le migliori pratiche di protezione dei dati. Gli sviluppatori dovrebbero evitare di utilizzare dati da fonti di terze parti come i broker di dati, poiché i loro dati potrebbero essere vecchi, distorti da bias, combinati in modo errato o avere altri difetti che potrebbero non essere stati corretti [125]. La raccolta attenta dei dati garantirebbe anche che un LMM non violi leggi sul copyright o sulla protezione dei dati, per le quali potrebbero esserci ripercussioni legali, tanto che alcuni LMM potrebbero essere considerati illegali [126].

Se vengono utilizzati fornitori di dati di terze parti, queste potrebbero, per esempio, essere certificate, al fine di costruire una fiducia e garantire la loro competenza e legittimità [127].
Tutti i dati utilizzati dagli sviluppatori per addestrare un LMM, raccolti direttamente o da terze parti, devono essere mantenuti aggiornati. Come notato in precedenza, alcuni dei principali modelli di IA non sono stati addestrati con dati aggiornati [38] e ciò può compromettere le prestazioni del modello nell'assistenza sanitaria e nella medicina, settori in cui le nuove prove che emergono di continuo influiscono significativamente sulle decisioni cliniche. I set di dati dovrebbero essere aggiornati e curati in modo che gli LMM siano appropriati e pertinenti ai contesti in cui vengono utilizzati.

Può essere difficile garantire che i dati siano adeguatamente trasparenti. Le aziende che pubblicano nuovi LMM sono diventate sempre meno trasparenti riguardo ai dati usati per addestrare i loro modelli. Una nota azienda leader nell'ambito dell'IA rilasciando un nuovo LMM ha dichiarato che: "dato il panorama competitivo e le implicazioni sulla sicurezza dei modelli su larga scala come GPT-

4, questo rapporto non contiene ulteriori dettagli riguardo all'architettura del modello (compresa la dimensione), all'hardware, al calcolo di addestramento, alla costruzione del dataset, al metodo di addestramento o simili" [128].

La mancanza di volontà di trasparenza riguardo ai dati è, tuttavia, incoerente con il principio etico dell'OMS che richiede la trasparenza, la "spiegabilità" e la comprensibilità. Gli sviluppatori dovrebbero essere trasparenti riguardo ai dati che hanno utilizzato per addestrare un modello in modo che le persone, comprese coloro che perfezionano ulteriormente l'LMM o coloro che utilizzano l'LMM per sviluppare un'applicazione sanitaria o gli utenti finali dell'LMM, siano consapevoli dei limiti e delle incompletezze del set di dati di addestramento.

Quando le aziende tecnologiche ricorrono a lavoratori dei Paesi a basso e a medio reddito per migliorare la qualità dei dati, per setacciare il contenuto onde eliminare materiale abusivo, violento o offensivo e per annotare i dati, devono garantire loro un salario dignitoso e un accesso ai servizi di salute mentale e ad altre forme di counselling; le aziende che sviluppano LMM devono introdurre misure di sicurezza per proteggere i lavoratori da qualsiasi distress. I governi a loro volta dovrebbero aggiornare le loro norme sul lavoro per estendere i benefici a tutti i lavoratori dei dati, promuovere un "gioco alla pari" tra le aziende e garantire che gli standard lavorativi siano mantenuti e migliorati nel tempo.

Progettazione etica e progettazione basata su valori: un approccio per integrare l'etica e lo standard dei diritti umani nello sviluppo delle tecnologie dell'IA è la "progettazione basata su valori", un paradigma che fonda la progettazione sui valori della dignità umana, della libertà, dell'uguaglianza e della solidarietà, considerandoli requisiti non funzionali ma basilari [1]. Diverse raccomandazioni nelle linee guida originali degli esperti dell'OMS per la progettazione delle tecnologie dell'IA, riguardano la "progettazione basata su valori" e meritano di essere riprese qui.

Le linee guida raccomandavano che la progettazione e lo sviluppo delle tecnologie dell'IA non fossero effettuati esclusivamente da ricercatori e ingegneri, ma che "potenziali utenti finali e tutte le parti interessate, dirette e indirette, dovrebbero essere coinvolti fin dalle prime fasi dello sviluppo dell'IA in una progettazione strutturata, inclusiva e trasparente e dovrebbero essere date opportunità di sollevare questioni etiche, esprimere preoccupazioni e fornire contributi per l'applicazione dell'IA in considerazione" [1]. Pertanto, nello sviluppo di modelli di base, le persone che potrebbero utilizzare o trarre beneficio dai modelli andrebbero coinvolte nello sviluppo iniziale. Una proposta è di introdurre i cosiddetti "collegi di supervisione umana", che faciliterebbero l'inclusione di rappresentanti dei pazienti nello sviluppo di un LMM destinato ad apportare benefici a un paziente o a un caregiver.[1] Nella progettazione di un LMM, nell'etichettatura dei dati e nei test andrebbero coinvolti professionisti sanitari, ricercatori, pazienti, persone comuni ed esponenti di gruppi vulnerabili. L'inclusività nella progettazione degli LMM potrebbe, per esempio, proteggere l'autonomia umana, perché la partecipazione dei fornitori di assistenza medica potrebbe prevenire o ridurre il bias di automazione. La

[1] Comunicazione di David Gruson, esperto WHO sull'etica e la governance dell'IA per la salute

progettazione inclusiva promuoverebbe inoltre il principio guida dell'OMS di garantire inclusività ed equità, specialmente se i team di progettazione includono punti di vista diversi per età, abilità, razza, etnia, sesso o identità di genere.

Le linee guida originali dell'OMS raccomandavano anche che "i progettisti e le altre parti interessate dovrebbero assicurarsi che i sistemi di IA siano costruiti per svolgere compiti ben definiti con l'accuratezza e l'affidabilità necessarie al fine di migliorare la capacità dei sistemi sanitari e promuovere gli interessi dei pazienti. I progettisti e le altre parti interessate dovrebbero anche essere in grado di prevedere e comprendere potenziali esiti secondari" [1]. Già prima di avviare lo sviluppo di un LMM, lo sviluppatore dovrebbe condurre una cosiddetta "pre mortem" [33] per considerare i "possibili fallimenti", in modo che il team di sviluppo possa ovviare in anticipo a questi fallimenti. Ciò consente agli sviluppatori di identificare rischi noti e non noti e di formulare alternative [33]. Un secondo suggerimento da parte di diversi sviluppatori di modelli di base per finalità generali è il "red teaming" [129], una valutazione di un modello o sistema che ne identifica le vulnerabilità in simulazioni del mondo reale che potrebbero portare a comportamenti indesiderati, come per esempio un LMM che fornisce un'opinione distorta a causa di bias, in modo che lo sviluppatore possa correggere il modello o il sistema per garantirne affidabilità e sicurezza. Una azienda ha presentato i suoi ultimi modelli di LMM alla conferenza DEFCON, una convention di hacker, in modo che "gli esperti potessero analizzare ulteriormente e testare le loro capacità" [130].

Le linee guida originali dell'OMS raccomandavano anche che "le procedure usate per la progettazione basata su valori dovrebbero essere informate e aggiornate in base al consenso, alle migliori pratiche (per esempio tecnologie che garantiscono la privacy), agli standard di etica progettuale e alle norme professionali in evoluzione" [1]. Una progettazione appropriata può limitare la divulgazione non autorizzata dei dati inseriti in un LMM o affrontare le preoccupazioni ambientali (carbonio e acqua) associate all'addestramento e all'uso degli LMM (vedi paragrafo successivo). Potrebbe anche garantire che gli utenti sappiano che i contenuti prodotti da un LMM sono generati da un sistema di IA e non da un essere umano, al fine di evitare di togliere all'essere umano la sua autorità epistemica. Tale notifica può ricordare agli utenti, alle comunità e alle società che mentre un LMM può produrre informazioni utili non può sostituire la produzione di conoscenza da parte degli esseri umani.

Progettazione nel rispetto delle questioni ambientali: come detto in precedenza, una preoccupazione importante riguardo agli LMM è la loro impronta di carbonio e idrica. Le aziende tecnologiche dovrebbero adottare tutte le possibili misure per ridurre il consumo di energia, per esempio migliorando l'efficienza energetica di un modello, e diverse grandi aziende stanno sperimentando tali approcci. Per esempio, una azienda ha sviluppato un LMM combinato con un database esterno che opera in modo più efficiente rispetto a un LMM [112]. Un'altra società sta sperimentando un LMM che si basa non su una rete neurale, ma distribuisce le sue variabili tra 64 reti neurali più piccole. Viene addestrato per utilizzare solo due reti neurali per completare ciascun compito, utilizzando così solo una piccola percentuale delle sue variabili per effettuare ciascuna inferenza [112].

Un altro mezzo per migliorare l'efficienza energetica è sviluppare LMM più piccoli che vengono addestrati su insiemi di dati più ridotti, che quindi richiedono meno energia per l'addestramento o il

funzionamento. Gli LMM più piccoli potrebbero non solo ridurre il consumo di energia, ma anche aprire opportunità per aziende o entità più piccole per sviluppare LMM migliorando l'accuratezza dei prodotti [59]. Gli LMM più piccoli potrebbero essere particolarmente utili per lo sviluppo di "LMM specializzati", come quelli destinati specificamente all'uso nell'assistenza sanitaria, nella ricerca scientifica e nella medicina. Alcuni LMM di questo tipo sono già stati introdotti, sviluppati anche da grandi aziende tecnologiche [59].

4.3 Leggi, politiche e investimenti del settore pubblico

Alcune leggi o politiche potrebbero essere applicate o scritte per ridurre o evitare i rischi durante la progettazione e lo sviluppo di modelli di base per finalità generali.
Inoltre, i governi potrebbero effettuare investimenti nel settore pubblico per promuovere o sostenere la progettazione e lo sviluppo etico di modelli di base per finalità generali.

Leggi e politiche che regolano l'uso dei dati: l'OMS sostiene l'applicazione e le azioni per assicurare il rispetto di norme, comprese le regole sulla protezione dei dati, che definiscono il modo in cui i dati devono essere utilizzati per addestrare gli LMM. Le leggi sulla protezione dei dati includono standard per la regolamentazione del trattamento dei dati che proteggono i diritti delle persone e stabiliscono, al contempo, obblighi per i responsabili del trattamento dei dati e per coloro che trattano i dati sia pubblici sia privati e includono sanzioni e risarcimenti in caso di azioni che violano i diritti stabiliti dalle leggi. Dato che leggi sulla protezione dei dati sono state adottate in oltre 150 Paesi, esse forniscono una solida base per lo sviluppo di tutte le tecnologie di IA, compresi gli LMM [1]. Un limite delle leggi sulla protezione dei dati è che la maggior parte di esse sono state emanate prima della comparsa dell'IA generativa e di altri tipi e usi dell'IA, e le autorità per la protezione dei dati potrebbero non essere disposte ad applicarle rigidamente, poiché le leggi in origine potrebbero non avere lo stesso intento [120].

Un requisito per la protezione dei dati che dovrebbe essere applicato, in particolare, per i dati sanitari utilizzati nell'addestramento degli LMM, è che i dati siano ottenuti e trattati in modo lecito. Ciò spesso richiede di ottenere un consenso informato pieno e consapevole da parte dell'interessato per l'uso dei suoi dati per lo scopo dichiarato. Ogni ulteriore trattamento deve avere una propria base giuridica, poiché non si può presumere che sia compatibile con lo scopo originario. Le aziende e le altre entità che hanno sviluppato e rilasciato gli LMM sono già sotto osservazione per il potenziale uso di dati ottenuti senza un consenso informato. La ricerca di LMM sempre più grandi, che richiedono set di dati sempre più ampi, può portare gli sviluppatori a ignorare i requisiti legali [83]. Ciò violerebbe anche il principio guida dell'OMS di proteggere l'autonomia umana. Pertanto, il gruppo di esperti dell'OMS ha raccomandato che i governi "dovrebbero dotarsi di leggi e regolamenti chiari sulla protezione dei dati", per l'uso dei dati sanitari e la protezione dei diritti individuali, compreso il diritto a un consenso informato pienamente consapevole.

Altre misure governative volte a sorvegliare e regolamentare la raccolta e l'uso dei dati per l'ad-

destramento degli LMM includono i regolamenti per l'IA generativa emanati dal governo cinese ed entrati in vigore nell'agosto 2023. L'Amministrazione del Cyberspazio cinese impone diversi obblighi, tra cui che:

1. **i fornitori adottino misure efficaci per evitare discriminazioni e pregiudizi nella selezione dei dati di addestramento**
2. **i fornitori utilizzino un'etichettatura chiara e valutino la qualità dell'etichettatura dei dati**
3. **gli sviluppatori adottino "misure efficaci" per raggiungere gli obiettivi di autenticità, accuratezza, obiettività ed eterogeneità dei dati [131].**

Non si prevede che i requisiti siano applicati in modo rigido dalle aziende, alle quali sarà richiesto solo di adottare misure efficaci per garantire un'adeguata qualità dei dati. Le misure si applicheranno solo alle aziende che offrono servizi al pubblico cinese [132].
Le disposizioni legislative relative ai dati potrebbero includere l'obbligo di descrivere le fonti di dati utilizzate per addestrare un modello di base e di utilizzare solo dati soggetti a una governance, inclusi l'idoneità, il rischio di bias e una mitigazione appropriata [133].
Di seguito sono descritte altre misure che i governi potrebbero adottare durante la progettazione e lo sviluppo di un LMM:

- *Profili di prodotto target:* **i governi e le agenzie internazionali potrebbero indicare profili di prodotti target per indicare le preferenze e le caratteristiche degli LMM destinati all'assistenza sanitaria e alla medicina, soprattutto se i governi prevedono di acquistare tali tecnologie per utilizzarle nei sistemi sanitari nazionali**
- *Standard e requisiti di progettazione e sviluppo:* **i governi potrebbero richiedere agli sviluppatori di garantire che la progettazione e lo sviluppo di un modello di base per finalità generali raggiunga determinati risultati durante il suo ciclo di vita. Potrebbero includere requisiti per la prevedibilità del modello e la sua interpretabilità, correggibilità, sicurezza e cybersicurezza [134]**
- *Programmi di pre-certificazione:* **le agenzie di regolamentazione potrebbero introdurre obblighi di legge e stabilire incentivi per richiedere e incoraggiare gli sviluppatori a identificare ed evitare rischi etici, quali la parzialità o l'indebolimento dell'autonomia umana, attraverso misure che includano programmi di pre-certificazione [1]. La precedente guida dell'OMS sull'etica dell'IA raccomanda che "le agenzie regolatorie governative dovrebbero fornire incentivi agli sviluppatori per identificare, monitorare e affrontare i problemi relativi alla sicurezza e ai diritti umani durante la progettazione e lo sviluppo del prodotto e dovrebbero integrare le linee guida rilevanti nei programmi di pre-certificazione" [1]**
- *Audit:* **i governi potrebbero introdurre audit sulle fasi iniziali di sviluppo dei modelli di base. Una proposta prevede tre tipi di audit: un "audit di governance" di un fornitore di LMM, un "audit degli LMM" e un "audit applicativo" dei prodotti e servizi a valle costruiti sugli LMM, che non si applicherebbero durante lo sviluppo di un LMM [121]. Gli audit potrebbero essere integrati nei requisiti per l'approvazione di un LMM destinato all'uso in ambito sanitario o medico. Affinché gli audit siano efficaci, la loro qualità deve essere valutata per garantire che soddisfino pienamente lo scopo previsto**
- *Impronta ambientale:* **i governi potrebbero richiedere agli sviluppatori di modelli di IA di base per finalità generali di affrontare la preoccupazione relativa alla loro impronta carbonica e idrica. Per esempio, i governi potrebbero richiedere agli sviluppatori di misurare il loro consu-**

mo energetico, di ridurre l'uso di energia durante l'addestramento [133] e di soddisfare standard ambientali non ancora definiti [134]

- ***Notifica che un contenuto prodotto da un LMM è "generato automaticamente":* i governi potrebbero richiedere agli sviluppatori di garantire che qualsiasi implementazione di un modello di base per finalità generali includa una notifica e un promemoria per gli utenti finali che il contenuto è stato generato da una macchina e non da un essere umano [133]**
- **I governi potrebbero anche prendere in considerazione la possibilità di prevedere od offrire incentivi per gli sviluppatori che registrino allo stadio di sviluppo iniziale algoritmi o sistemi di IA da utilizzare nell'assistenza sanitaria o in ambito medico. La registrazione iniziale potrebbe incoraggiare la pubblicazione di risultati negativi, prevenire il bias di pubblicazione o un'interpretazione troppo ottimistica dei risultati e facilitare l'integrazione delle conoscenze a beneficio dei pazienti.**

L'infrastruttura pubblica per sviluppare gli LMM nell'interesse pubblico: con la proliferazione degli usi degli LMM per la salute, lo sviluppo degli LMM che aderiscono ai principi etici potrebbe essere incoraggiato dalla fornitura di infrastrutture pubbliche o senza scopo di lucro, tra cui potenza di calcolo e set di dati pubblici. Tale infrastruttura, accessibile agli sviluppatori del settore pubblico, privato e no-profit, potrebbe richiedere agli utenti di aderire a principi e valori etici in cambio dell'accesso. Potrebbe inoltre contribuire a evitare il controllo esclusivo di un LMM da parte di uno sviluppatore e a "determinare un campo di gioco equo" tra le aziende più grandi e gli sviluppatori che non hanno accesso a grandi infrastrutture e risorse.

I governi, soggetti a una supervisione indipendente, potrebbero costruire un'infrastruttura che venga poi utilizzata dagli sviluppatori per costruire gli LMM per l'assistenza sanitaria e la medicina. Per esempio, un gruppo internazionale di mille volontari accademici, l'azienda Hugging Face e altri, con fondi del governo francese, hanno addestrato un LMM chiamato BLOOM con 175 miliardi di parametri, che ha richiesto 7 milioni di dollari per il tempo di calcolo [112].

Gli sforzi per creare un campo di gioco equo sono applicabili anche al mondo accademico e al suo svantaggio in termini di risorse. La piattaforma nazionale di calcolo per la ricerca avanzata del governo canadese è al servizio del settore accademico del Paese, il governo cinese ha approvato un sistema di rete nazionale di potenza di calcolo per consentire agli accademici e ad altri soggetti di accedere ai dati e alla potenza di calcolo e negli Stati Uniti la task force per la ricerca nazionale sull'IA ha "proposto la creazione di un cloud informatico per la ricerca e un dataset pubblico" [101]. In Europa ci sono anche appelli della società civile ai governi per svolgere un ruolo più incisivo nella costruzione dei cosiddetti "grandi modelli generativi europei", per i quali i governi fornirebbero l'elaborazione, l'infrastruttura di dati, il supporto alla scienza e alla ricerca specifici per l'IA [135].

4.4 LMM open source

Il ruolo degli LMM open source nell'integrazione dei principi etici e nell'affrontare i rischi noti è incerto. In generale, la trasparenza e la partecipazione possono essere aumentate utilizzando software open source per la progettazione di una tecnologia di IA o rendendo il codice sorgente del software pubblicamente disponibile [1]. Il software open source è aperto sia ai contributi sia ai feedback e ciò consente agli utenti di capire come funziona il sistema, di identificare i potenziali problemi e di estendere e adattare il software [1]. Gli LMM open source possono essere un'opportunità per gestire alcune preoccupazioni sull'uso degli LMM in ambito sanitario. Poiché i modelli open source non sono né proprietari né chiusi, consentono alle aziende e alle entità più piccole, come le istituzioni no-profit, di progettare LMM a costi inferiori [136]. Gli LMM costruiti su modelli open source possono essere esaminati, poiché il codice e i dati sono disponibili per la revisione. Il coinvolgimento e il controllo da parte di una comunità di utenti contribuiscono a garantire la solidità dei modelli open source nel lungo periodo [136]. Tuttavia, gli LMM open source potrebbero non sopravvivere se le grandi aziende tecnologiche che in passato hanno reso disponibili i loro modelli decidessero di non continuare a farlo [10]. Lo sviluppo della maggior parte degli LMM open source si è basato su un LMM rilasciato su base limitata da Meta (ex Facebook) [10]. Da quando l'LMM e i suoi pesi sono trapelati [137], l'azienda ha dichiarato il proprio impegno a favore di approcci open source, osservando che l'apertura "porta a prodotti migliori, a un'innovazione più rapida e a un mercato fiorente, che va a vantaggio di [Meta] come di molti altri… in definitiva, l'apertura è il miglior antidoto alle paure che circondano l'IA" [130]. Osservatori indipendenti hanno tuttavia notato che, sebbene Meta abbia reso disponibile il suo LMM su base non commerciale, le sue condizioni d'uso includono restrizioni, dunque non sta fornendo il suo LMM in modo coerente con i principi dell'open source [138, 139].

Ulteriori requisiti per l'utilizzo di modelli open source al fine di monitorarne le prestazioni e i risultati saranno difficili da gestire per gli sviluppatori; tuttavia, il vantaggio di tali modelli non può sostituire la necessità di regolamentare ed evitare i rischi, come per esempio i problemi di sicurezza associati all'uso di modelli open source [140]. I modelli open source possono essere esposti a un uso improprio [141] e possono essere attaccati per sfruttare tale vulnerabilità [142]. Un gruppo di ricercatori ha recentemente scoperto che i metodi testati su sistemi di IA open source hanno aggirato le misure di sicurezza e di tutela dell'IA e potrebbero anche aggirare le salvaguardie dei cosiddetti sistemi chiusi [143]. In definitiva, i modelli open source si basano sulle stesse tecnologie a "scatola nera" utilizzate in altri LMM.

Un modo per incoraggiare gli LMM open source sarebbe di richiedere ai governi che i modelli di base costruiti con finanziamenti pubblici o con la proprietà intellettuale del governo siano ampiamente accessibili, nello stesso modo in cui i governi hanno richiesto l'accesso aperto ai risultati della ricerca finanziata dal governo. I governi potrebbero anche incoraggiare la ricerca e lo sviluppo open source nelle strutture pubbliche, compresi i modelli di nuova generazione, in condizioni controllate e con una supervisione pubblica. La supervisione e la partecipazione pubblica potrebbero essere migliori della nuova realtà nella quale il modello trapelato di Meta permette a chiunque di "scaricarlo ed eseguirlo su un MacBook M2" [144].

Raccomandazioni:

- **Gli sviluppatori che progettano un LMM che deve o potrebbe essere utilizzato nell'assistenza sanitaria, nella ricerca scientifica o in ambito medico, devono considerare di avere una certificazione etica e di formare i programmatori. Questo porterebbe gli sviluppatori di IA in linea con i requisiti della professione medica e aumenterebbe la fiducia nei loro prodotti e servizi**
- **Indipendentemente dalle dimensioni del set di dati, gli sviluppatori dovrebbero intraprendere "valutazioni d'impatto sulla protezione dei dati" prima di trattare tali dati. Ciò richiede agli sviluppatori di valutare il rischio che le operazioni di trattamento dei dati vadano contro i diritti e la libertà delle persone e il suo impatto sulla protezione dei dati personali**
- **Gli sviluppatori dovrebbero addestrare gli LMM sui dati raccolti secondo le migliori pratiche e il rispetto delle leggi sulla protezione dei dati**
- **Tutti gli insiemi di dati utilizzati per addestrare gli LMM, raccolti sia direttamente sia tramite terzi dagli sviluppatori, devono essere aggiornati e adeguati ai contesti in cui il sistema può essere utilizzato**
- **Gli sviluppatori dovrebbero essere trasparenti sui dati utilizzati per addestrare un modello, così che gli utilizzatori, compreso chi si occupa dell'affinamento degli LMM, chi usa l'LMM per sviluppare un'applicazione di uso sanitario o chi usa l'LMM direttamente, siano consapevoli delle carenze e dell'incompletezza del set di dati di addestramento**
- **Gli sviluppatori dovrebbero pagare i lavoratori dei dati con un salario adeguato indipendentemente dal Paese in cui operano e fornire loro servizi di salute mentale e altre forme di counselling. Le aziende dovrebbero inoltre introdurre misure di salvaguardia per proteggere i lavoratori da eventuali distress. I governi dovrebbero aggiornare le norme sul lavoro per estendere tali benefici a tutti i lavoratori dei dati, per promuovere una "parità di condizioni" tra le aziende e per garantire che tali norme sul lavoro siano mantenute e migliorate nel tempo**
- **Gli sviluppatori dovrebbero garantire che gli LMM non siano progettati solo da ricercatori e ingegneri. I potenziali utenti e tutte le parti interessate dirette e indirette, compresi i fornitori di servizi medici, altri ricercatori, gli operatori sanitari e i pazienti, dovrebbero essere coinvolti fin dalle prime fasi dello sviluppo dell'IA in una progettazione strutturata, inclusiva e trasparente e avere l'opportunità di sollevare questioni etiche, esprimere preoccupazioni e fornire input per l'applicazione dell'IA in esame. Tale contributo potrebbe essere fornito attraverso "collegi di supervisione umana"**
- **Gli sviluppatori devono assicurare che gli LMM siano progettati per svolgere compiti ben definiti con l'accuratezza e l'affidabilità necessarie per migliorare la capacità dei sistemi sanitari e promuovere gli interessi dei pazienti. Gli sviluppatori dovrebbero anche essere in grado di prevedere e comprendere i potenziali esiti secondari. Le tecniche per soddisfare tali requisiti includono le "pre-mortem" e il "red teaming"**
- **Le procedure utilizzate dagli sviluppatori per la "progettazione basata su valori" dovrebbero essere informate e aggiornate in base al consenso, alle migliori pratiche (per esempio, tecnologie per preservare la privacy), agli standard di etica della progettazione e alle norme professionali in evoluzione, compresa la divulgazione del fatto che i contenuti prodotti da un LMM sono generati da un sistema di IA**
- **Gli sviluppatori dovrebbero adottare tutte le misure possibili per ridurre il consumo energetico (per esempio migliorando l'efficienza energetica di un modello)**

- **I governi dovrebbero avere leggi e regolamenti forti sulla protezione dei dati in ambito sanitario che vengano fatti rispettare nello sviluppo degli LMM. Le leggi devono proteggere efficacemente i diritti delle persone e fornire loro gli strumenti necessari per tutelarli, compreso il diritto a un consenso informato pienamente consapevole. Probabilmente saranno necessari ulteriori strumenti per i dati raccolti ed elaborati per l'uso degli LMM nell'assistenza sanitaria**
- **I governi e le agenzie internazionali, come l'OMS, dovrebbero indicare dei "profili di prodotto target" per delineare le preferenze e le caratteristiche degli LMM destinati all'assistenza sanitaria e alla medicina, soprattutto se i governi prevedono un eventuale acquisto di tali strumenti da utilizzare nei sistemi sanitari nazionali**
- **I governi dovrebbero richiedere agli sviluppatori di garantire che la progettazione e lo sviluppo di un modello di base per finalità generali raggiungano determinati risultati durante il ciclo di vita del prodotto. Questi potrebbero includere requisiti per la prevedibilità del modello e la sua interpretabilità, correggibilità, sicurezza e cybersicurezza**
- **Le agenzie regolatorie dovrebbero introdurre obblighi legali e stabilire incentivi, come i programmi di pre-certificazione, per richiedere e incoraggiare gli sviluppatori a identificare ed evitare i rischi etici, tra cui la parzialità o la compromissione dell'autonomia**
- **I governi dovrebbero introdurre audit sulle fasi iniziali di sviluppo dei modelli di base**
- **I governi dovrebbero richiedere agli sviluppatori di modelli di base per finalità generali di affrontare le preoccupazioni relative all'impronta carbonica e idrica**
- **I governi dovrebbero richiedere agli sviluppatori di garantire che, in qualsiasi utilizzo di un modello di base per finalità generali, gli utenti siano avvisati che il contenuto è stato generato da una macchina e non da un essere umano**
- **I governi dovrebbero prendere in considerazione la possibilità di richiedere o creare incentivi per gli sviluppatori affinché registrino in fase iniziale gli algoritmi o i sistemi di IA utilizzati nell'assistenza sanitaria e nella medicina. La registrazione iniziale potrebbe incoraggiare la pubblicazione dei risultati negativi, prevenire il bias di pubblicazione o l'interpretazione troppo ottimistica dei risultati e facilitare l'inclusione di conoscenze utili per i pazienti**
- **I governi dovrebbero investire o fornire infrastrutture, pubbliche o senza scopo di lucro, compresa la potenza di calcolo e gli insiemi di dati pubblici, accessibili agli sviluppatori del settore pubblico, privato e senza scopo di lucro, che richiedano agli utenti di aderire a principi e valori etici in cambio dell'accesso**
- **I governi dovrebbero incoraggiare lo sviluppo degli LMM open source richiedendo che i modelli di base costruiti con finanziamenti pubblici o con proprietà intellettuale del governo siano ampiamente accessibili, nello stesso modo in cui i governi hanno richiesto l'accesso aperto ai risultati della ricerca finanziata dal governo. I governi dovrebbero sostenere la ricerca e lo sviluppo open source in strutture pubbliche, compresi i modelli di nuova generazione, in condizioni controllate, con una supervisione pubblica.**

5. Disposizioni relative ai modelli di base per finalità generali

Gli usi dei modelli di base per finalità generali dipendono da come un utente interroga l'LMM per generare risposte relative all'assistenza sanitaria o da come un fornitore sia autorizzato dallo sviluppatore a integrare l'LMM in un'applicazione, un prodotto o un servizio per l'assistenza sanitaria. In entrambi i casi, si introducono dei rischi nuovi che devono essere gestiti dagli sviluppatori, dai fornitori o da entrambi. I governi sono responsabili di valutare e regolamentare l'uso di tali tecnologie prima del loro rilascio.

5.1 Rischi che devono essere affrontati quando si fornisce un servizio o un'applicazione sanitaria con un modello di base per finalità generali

È probabile che ci sia disaccordo sul fatto che dovrebbero essere valutati e approvati sia i modelli di base per finalità generali sia le applicazioni quando vengono utilizzate a fini sanitari o direttamente dall'utente. Alcune delle più grandi aziende tecnologiche hanno fatto pressioni sui funzionari governativi (per esempio nell'Unione Europea) per non occuparsi della valutazione degli LMM e concentrarsi invece sulla supervisione delle applicazioni che potrebbero essere utilizzate in modi che un governo potrebbe considerare rischiosi [145]. Ciò riguarderebbe sia i fornitori che generano e commercializzano applicazioni sanitarie che includono il modello di base, sia gli utenti, come un fornitore o un paziente, che scelgono di utilizzare direttamente o indirettamente l'LMM tramite un sistema di IA. Le aziende sostengono che la supervisione del modello di base per finalità generali "porrebbe completamente il carico" sugli sviluppatori mentre anche altri attori della catena del valore dovrebbero assumersi responsabilità [145].

Se può non essere opportuno ritenere responsabile lo sviluppatore di un modello di base per finalità generali per tutti gli utilizzi dell'LMM che possono essere fatti, sarebbe altrettanto inappropriato porre il carico esclusivamente sui fornitori, i gestori o gli utenti, poiché non sono stati coinvolti nello sviluppo del modello e potrebbero non comprenderne le limitazioni e rischi associati. Ciò consentirebbe inoltre agli sviluppatori di modelli di base per finalità generali, nonostante il loro potere, le loro risorse, la loro supervisione e comprensione degli LMM, di sfuggire alle responsabilità, aprendo un "enorme buco" nei tentativi di governare le tecnologie di IA per la salute [145].

Uno sviluppatore potrebbe cercare di evitare l'uso di un proprio LMM per scopi sanitari. Se uno sviluppatore non desidera che un LMM venga utilizzato per scopi sanitari o medici (specialmente in clinica), potrebbe scoraggiarne l'utilizzo sia impedendo agli enti che sviluppano applicazioni per la salute o la medicina di avere una licenza di utilizzo dell'LMM su un'interfaccia di programmazione

delle applicazioni (API), sia, se l'LMM viene utilizzato direttamente da un utente (fornitore o paziente) per scopi sanitari, mediante il blocco delle domande e delle richieste in questo ambito oppure fornendo un chiaro avvertimento a qualsiasi risposta che includa informazioni sanitarie o mediche e indirizzando gli utenti a informazioni o servizi che possono fornire un'assistenza appropriata.
Se tali misure non vengono adottate o se lo sviluppatore intende che il suo LMM venga usato nell'assistenza sanitaria, direttamente o indirettamente attraverso un fornitore, lo sviluppatore avrà specifiche responsabilità, di cui deve rispondere. Inoltre, sia gli sviluppatori sia i fornitori hanno ulteriori obblighi per gestire i rischi associati all'uso degli LMM nell'assistenza sanitaria.

Le responsabilità, che vengono riportate di seguito, sono definite da leggi, politiche e regolamenti, poiché sono i governi che devono determinare se un sistema basato sull'IA dovrebbe essere permesso per l'uso nell'assistenza sanitaria. Sia gli sviluppatori sia i fornitori devono inoltre rispondere a responsabilità reciproche se un LMM è utilizzato nell'assistenza sanitaria. Tali responsabilità potrebbero essere definite dai governi o negoziate tra le due parti attraverso un contratto qualora le leggi non prevedano ancora esplicitamente i doveri.
I principali rischi che devono essere gestiti prima dell'implementazione includono i bias a livello di sistema, le informazioni false o le "allucinazioni" per usi sanitari, la privacy dei dati inseriti in un LMM, la manipolazione e il bias di automazione.

5.2 Misure che i governi possono introdurre per gestire i rischi e i principi etici da osservare

La velocità di sviluppo degli LMM e delle applicazioni che includono un LMM richiede che i governi sviluppino rapidamente regole e criteri specifici per l'uso di questi algoritmi di IA nei sistemi sanitari e per altri scopi medici e scientifici. L'approccio dovrebbe consistere nella valutazione e nell'approvazione delle tecnologie basate sull'IA destinate all'uso nell'assistenza sanitaria o nella medicina da parte di un'agenzia regolatoria: potrebbe essere un'agenzia che si occupa dei dispositivi medici o dei farmaci, ma i governi potrebbero istituire anche una nuova agenzia a tal fine. È questa una sfida per i Paesi a basso e a medio reddito perché le loro agenzie regolatorie sono già sottodimensionate e sopraffatte dalle attività per la regolamentazione farmaceutica.
Il governo di un Paese ad alto reddito come gli Stati Uniti ha concordato con le maggiori aziende tecnologiche che i modelli di base saranno sottoposti a una valutazione pubblica volontaria, con divulgazione dei risultati per fornire informazioni al pubblico e ai ricercatori riguardo ai modelli e per incoraggiare le aziende a correggere eventuali errori [146]; comunque, un approccio volontario non è probabilmente sufficiente né sostenibile.

La valutazione degli LMM e delle applicazioni non dovrebbe riguardare solo i sistemi o gli algoritmi di IA utilizzati nel sistema sanitario, poiché ci sono anche rischi significativi associati all'uso di LMM e applicazioni in una zona grigia di intersezione tra applicazioni per la clinica e per il "benessere della persona". Vista la rapida proliferazione degli LMM, i governi, almeno inizialmente, dovrebbero identificare le applicazioni, stabilire standard e regolamenti comuni e impedire che le applicazioni che non soddisfano gli standard e i regolamenti vengano distribuite al pubblico.

Gli sviluppatori e i fornitori dovrebbero fornire l'onere della prova, quando richiesto, per dimostrare che la loro tecnologia di IA destinata all'uso nella sanità soddisfa i requisiti minimi stabiliti dalla legge o da un policy governativa. Non bisogna dare per scontato che visto che si conoscono i rischi e le sfide associati agli LMM, gli algoritmi di IA e le applicazioni con un LMM siano sicuri ed efficaci o che siano superiori ad altri approcci, basati sull'IA o meno, già ampiamente utilizzati.

Diverse leggi, policy e requisiti che potrebbero essere applicati all'uso degli LMM nella sanità e nella medicina sono descritti di seguito.

Requisiti di divulgazione (trasparenza): una regolamentazione appropriata richiede non solo che i governi abbiano la capacità e la discrezionalità per decidere che cosa valutare e approvare per l'uso, ma anche che abbiano le informazioni adeguate per condurre tale valutazione. La divulgazione è necessaria sia per regolare adeguatamente una tecnologia di IA sia per garantire che altri attori della catena del valore possano utilizzare la tecnologia in modo sicuro. Per esempio, a meno che uno sviluppatore non divulghi le prestazioni di un modello di base per finalità generali (come la sua propensione all'"allucinazione"), un fornitore potrebbe non avere le informazioni necessarie per ottimizzare il modello o evitare di commercializzare la tecnologia. Tali forme di divulgazione da parte di un fornitore o uno sviluppatore possono anche aiutare gli utenti fornitori di servizi per l'assistenza medica nella decisione se usare un LMM che potrebbe fornire informazioni errate o nell'esaminare più attentamente i risultati.
La divulgazione e la trasparenza sono principi guida dell'OMS, così come misure per migliorare la "spiegabilità" e l'intelligibilità di un sistema basato sull'IA, e dovrebbero essere richieste nella valutazione di un modello o applicazione di base per finalità generali. Le linee guida dell'OMS "Ethics and governance of artificial intelligence for health" [1] hanno raccomandato che "le agenzie regolatorie governative dovrebbero richiedere la trasparenza su determinati aspetti di una tecnologia di IA, tenendo presenti i diritti di proprietà, per migliorare la supervisione e la valutazione della sicurezza e dell'efficacia. Ciò può includere il codice sorgente di una tecnologia di IA, gli input di dati e l'approccio analitico". Nuove forme di divulgazione importanti per gli LMM possono riguardare le loro prestazioni nei test interni e la loro impronta di carbonio e idrica. Potrebbero essere necessari standard anche per i "pesi aperti", che consentono ai regolatori, ad altri sviluppatori, alla società civile e ai fornitori di comprendere gli output dell'addestramento di un algoritmo o le conoscenze che un LMM ha acquisito durante il suo addestramento [147, 148].
Diverse forme di divulgazione potrebbero aiutare un fornitore, un utente o un ente regolatore, tra cui la descrizione delle capacità e dei limiti di un modello di base, la valutazione del modello secondo riferimenti standard pubblici o del settore e la descrizione dei risultati dei test interni ed esterni del modello e della sua ottimizzazione [134]. La dichiarazione dei rischi che possono essere associati a un LMM o a un'applicazione va divulgata chiaramente, tanto che un ricercatore ha paragonato queste informazioni a quelle di un'etichetta nutrizionale [129].

Leggi sulla protezione dei dati: lo sviluppo degli LMM e il modo in cui gli sviluppatori gestiscono i dati necessari per addestrare un LMM potrebbero violare le leggi sulla protezione dei dati. Un pro-

blema ulteriore è che i dati inseriti in un LMM o un'applicazione mirati a un risultato specifico possono includere informazioni personali o sensibili che potrebbero essere diffuse accidentalmente o attraverso le interrogazioni fatte all'LMM. La potenziale diffusione di questi dati è la ragione per cui molte grandi aziende, comprese le aziende tecnologiche che stanno sviluppando e commercializzando LMM, vietano ai propri dipendenti l'uso di questi algoritmi [149].
La diffusione dei dati viola la responsabilità dello sviluppatore nel proteggere l'autonomia. Gli sviluppatori possono anche violare le leggi sulla protezione dei dati se questi sono conservati più a lungo del consentito secondo i requisiti di minimizzazione dei dati. Un produttore consente agli utenti di scegliere se escludere i contenuti da loro forniti per perfezionare le prestazioni del suo chatbot [150]. I governi che consentono l'uso di LMM dovrebbero chiedere garanzia che questi rispettino le regole sulla protezione dei dati inseriti nell'LMM. Le norme del governo cinese per gli LMM includono tale requisito, anche se la protezione si applica solo agli utenti in Cina [151].

Valutazione dei modelli e/o delle applicazioni di base per finalità generali utilizzate nella sanità: la legge sui diritti umani rispetto a leggi basate sul rischio: sono stati approvati diversi impianti di legge per valutare e regolamentare le tecnologie di IA. Una questione è se le tecnologie di IA debbano soddisfare gli obblighi dei diritti umani ("diritti fondamentali" secondo l'Unione Europea) o se debba essere utilizzato un approccio diverso per valutare le tecnologie di IA in un contesto basato sul rischio [152]. L'Unione Europea, nell'ambito dell'*AI Act*, ha adottato un contesto basato sul rischio, che potrebbe contribuire a identificare i requisiti o l'onere della prova che devono essere forniti per una tecnologia, onere della prova che aumenta con il livello di rischio della tecnologia.
Tutti i sistemi e gli strumenti di IA utilizzati nella sanità e nella medicina dovrebbero rispettare i principi etici e gli standard dei diritti umani che riguardano, per esempio, la dignità, l'autonomia o la privacy di una persona. Questi includono i modelli di base per finalità generali. I principi etici e dei diritti umani non sono negoziabili e devono essere rispettati, indipendentemente dal rischio associato a una tecnologia di IA o dal beneficio che può conferire. Il fatto che un algoritmo di IA sia considerato "a basso rischio" non lo esenta da un'analisi, e uno sviluppatore o un fornitore dovrebbero sempre garantire che l'algoritmo rispetti i diritti umani e i principi etici. Una valutazione del rispetto dei diritti umani può essere condotta per determinare se un LMM o un'applicazione aderisca a tali impegni e possa quindi essere utilizzato in modo sicuro. La guida dell'OMS sull'etica e la governance dell'IA per la salute [1] raccomanda che "i governi dovrebbero emanare leggi e avere policy che impongano alle agenzie governative e alle aziende di condurre valutazioni di impatto delle tecnologie di IA per quanto concerne l'etica, i diritti umani, la sicurezza e la protezione dei dati, durante tutto il ciclo di vita di un sistema di IA". La guida ha inoltre osservato che "le valutazioni di impatto dovrebbero essere verificate da un terzo attore indipendente prima e dopo l'introduzione di una tecnologia di intelligenza artificiale e rese pubbliche" [1]. I risultati delle valutazioni di impatto vanno quindi rese pubbliche, tenendo conto degli aspetti di proprietà e di informazioni sensibili, e dovrebbero essere disponibili a tutti coloro che potrebbero essere interessati. Come per gli audit (vedi prima), è necessario esaminare attentamente le valutazioni di impatto, soprattutto se vengono condotte da soggetti terzi che offrono strumenti o servizi, per garantire che essi siano di qualità e rigore adeguati.

Le valutazioni di impatto possono rivelare, per esempio, se una tecnologia di IA introduca pregiudizi a livello di sistema, rischi per la privacy degli utenti che condividono dati personali o manipolazione degli utenti stessi. I rischi per la privacy vanno affrontati mediante la collaborazione tra fornitori e sviluppatori di LMM. Una sperimentazione su un LMM di questo tipo è in corso tra un'azienda tecnologica e un sistema ospedaliero negli Stati Uniti, ma il successo del progetto è considerato improbabile perché i dati sanitari non possono essere completamente resi anonimi [154].

Le valutazioni di impatto possono anche garantire che l'uso di un modello di base per finalità generali o di un'applicazione informi gli utenti al fine di evitare un processo decisionale automatizzato ripetuto in cui un utente riceve informazioni false o disinformazioni o un operatore sanitario o un paziente fanno affidamento in modo acritico sull'output di un LMM, creandosi così un bias di automazione. I governi potrebbero invece scegliere di utilizzare un contesto di leggi basato sul rischio per gli LMM utilizzati in ambito della medicina e della sanità. Per le funzioni considerate a rischio più elevato, come la fornitura di una prescrizione o un consiglio sulla salute mentale a una persona affetta da depressione grave o l'uso di una tecnologia di IA da parte di una popolazione vulnerabile o emarginata, l'onere della prova deve essere più elevato. La preoccupazione è che, nel caso in cui la legge scelga un approccio basato sul rischio, questo venga considerato come sufficiente o utilizzato come sostituto di un approccio basato sui diritti umani [153]. Un contesto basato sul rischio potrebbe escludere alcuni LMM o applicazioni dalla valutazione, cosa che potrebbe sembrare di scarso peso ma che alla fine potrebbe arrecare danni. Ulteriori domande sono se la valutazione regolatoria dell'IA debba applicarsi ai modelli di base indipendentemente dal loro utilizzo, se la valutazione debba applicarsi solo agli LMM più grandi e più utilizzati ("modelli di base sistematici"[1]) e in quali casi tali valutazioni debbano applicarsi ai fornitori.

La guida non include raccomandazioni riguardo al fatto se tutti i modelli di base, indipendentemente dal modo in cui vengano utilizzati, debbano essere soggetti a un processo di valutazione basato sul rischio e/o sui diritti. La guida inoltre non si esprime riguardo al fatto se una valutazione regolatoria dell'IA dei modelli di base per finalità generali debba applicarsi solo ai più grandi (sistemici). Il gruppo di esperti ha sollevato preoccupazioni rispetto a una valutazione costruita per essere applicabile a tutti i modelli, a prescindere dalla grandezza, poiché potrebbe sancire il dominio delle aziende più grandi, in quanto gli standard diverrebbero tali che solo quelle aziende potrebbero veramente rispettarli o comunque gli standard si adatterebbero meglio al loro modello di business e ai loro obiettivi [155]. Questa preoccupazione ha attirato l'attenzione delle autorità garanti della concorrenza, le quali hanno scoperto che chi si è mosso per primo può aver usato "metodi di concorrenza sleali per consolidare il suo attuale potere o utilizzare tale potere per acquisire il controllo su un nuovo mercato dell'IA generativa" [141]. Le autorità garanti della concorrenza avranno la responsabilità di esercitare un maggiore controllo sull'uso degli LMM, sebbene si concentrino sulle pratiche delle imprese che sviluppano gli LMM [156].

Anche i fornitori dovrebbero essere soggetti a processi di valutazione regolatoria sull'IA, dato che il loro utilizzo di un LMM potrebbe cambiarne lo scopo e la funzione rispetto a quelli definiti dallo

[1] Presentazione di Kai Zenner, European Parliament, Head of Office for Axel Voss, Conference on AI for Good, 5 giugno 2023

sviluppatore. Pertanto, se un modello di base per finalità generali è adattato all'uso in sanità o in medicina da un fornitore, a cui lo sviluppatore dà il proprio assenso, sia lo sviluppatore sia il fornitore dovrebbero rispettare i requisiti di utilizzo degli LMM in sanità e in medicina. L'onere normativo imposto ai fornitori dovrebbe essere maggiore se l'uso che fanno di un prodotto o applicazione diverge in maniera sostanziale dal modello di base o lo cambia in modi che vanno oltre il controllo dello sviluppatore.

Normativa sui dispositivi medici: un governo potrebbe stabilire che un modello di base per finalità generali o un'applicazione venga proposto come dispositivo medico. Dato che al momento vi sono pochi elementi per definire quando un LLM è qualificabile come un dispositivo medico, un ente regolatore ha affermato che "gli LMM diretti solo a scopi generali e per i quali gli sviluppatori non dichiarino che il proprio software sia utilizzabile a fini medici difficilmente sono qualificabili come dispositivi medici" [157]. Il regolatore ha fatto, tuttavia, inoltre notare come "gli LMM che sono sviluppati, adattati, modificati o diretti a fini specificamente medici sarebbero invece verosimilmente qualificabili come dispositivi medici". In più, qualora uno sviluppatore affermi che il proprio LMM possa essere utilizzato a scopi medici, anche questo avvalorerebbe la qualifica di dispositivo medico" [157].
È verosimile che i chatbot basati su LMM che forniscono consigli medici siano caratterizzati come dispositivi medici secondo gli attuali standard normativi dell'Unione Europea e degli Stati Uniti [158]. Le linee guida dell'OMS sull'etica e la governance dell'IA per la salute [1] raccomandano che: "le autorità di regolamentazione governative dovrebbero richiedere che le performance di un sistema di IA siano testate e che si ottengano prove solide mediante valutazione prospettica in studi controllati e randomizzati e non semplicemente dal confronto del sistema con set di dati pregressi esistenti in un laboratorio".

Se un LMM o un'applicazione devono essere regolamentati come dispositivo medico, lo sviluppatore e/o il fornitore dovrebbero sostenere l'onere della prova fornendo evidenza che il dispositivo funzioni come commercializzato e che soddisfi i requisiti delle leggi nazionali. Ciò potrebbe includere vari requisiti, come il rispetto degli obblighi etici legati ai pregiudizi e alla privacy. Nuovi regolamenti proposti sulle tecnologie di IA per i dispositivi medici nell'Unione Europea e negli Stati Uniti integreranno probabilmente i principi etici legati all'uso dell'IA per la salute, inclusi la "spiegabilità", il controllo dei bias e la trasparenza. È improbabile che gli attuali chatbot che includono LMM possano soddisfare tali standard [158].

Gli LMM per il supporto alle decisioni cliniche sono già utilizzati a livello sperimentale. Sebbene questi LMM includano dichiarazioni di non responsabilità, esse non ovviano all'applicazione delle leggi sui dispositivi medici che impongono che "tali sperimentazioni debbano avere luogo solo nell'ambito di una sperimentazione clinica autorizzata, con controlli appropriati per proteggere i pazienti e per produrre risultati clinicamente rilevanti" [158]. I governi potrebbero valutare usi sperimentali controllati di tali LMM con normative mirate, che consentirebbero di eseguire test in ambienti dal vivo in condizioni cliniche reali ma con misure di salvaguardia e supervisione per proteggere il si-

stemi sanitario da rischi o conseguenze indesiderate. Tale utilizzo, tuttavia, potrebbe essere appropriato solo nei Paesi in cui i nuovi prodotti e servizi sanitari e le relative specifiche sono soggetti a regolamenti formali e di protezione dei dati [1].

Legge sulla tutela dei consumatori: i governi dovrebbero promulgare leggi sulla tutela dei consumatori per garantire che eventuali conseguenze negative degli LMM e delle applicazioni non colpiscano gli utenti e i pazienti. Le leggi sulla tutela dei consumatori potrebbero essere applicate, per esempio, per prevenire pratiche equivalenti a una manipolazione dell'utente [159]. Negli Stati Uniti, diversi dipartimenti governativi e agenzie stanno applicando leggi sulla tutela dei consumatori e altri regolamenti per prevenire discriminazioni e bias nei sistemi automatizzati [159]. Tali leggi possono consentire ai governi di richiedere agli enti che cercano di commercializzare tali tecnologie di affrontare le cause di eventuali conseguenze negative e di proteggere i pazienti e le loro famiglie da qualsiasi danno, presente o futuro [93]. Le leggi sulla tutela dei consumatori, o altre norme, potrebbero essere utilizzate per richiedere che gli LMM e le applicazioni siano limitati nell'uso di un linguaggio che potrebbe lasciar intendere o fuorviare un utente finale nell'attribuire qualità umane a un LMM. Tali leggi potrebbero quindi limitare l'uso degli LMM e applicazioni o impedire di usare nelle risposte parole come "Io penso", "Io suppongo" o "Io suggerisco".

Raccomandazioni:

- **Se o lo consentono le risorse, i governi dovrebbero incaricare un'agenzia regolatoria, già esistente o nuova, di valutare e approvare gli LMM e le applicazioni destinate all'uso in sanità e in medicina**
- **Diversi aspetti di un LMM e le sue applicazioni dovrebbero essere trasparenti per consentire la supervisione e la valutazione della loro sicurezza ed efficacia da parte delle autorità regolatorie. Ciò potrebbe includere il codice sorgente, gli input dei dati, i pesi del modello e l'approccio analitico. Ulteriori forme di trasparenza che un governo deve prendere in considerazione sono le performance di un LMM (o delle applicazioni) nei processi di test interni e la sua impronta di carbonio e idrica**
- **I governi dovrebbero garantire che le norme sulla protezione dei dati si applichino ai dati inseriti in un LMM o nelle applicazioni**
- **Le leggi, le policy e i regolamenti governativi dovrebbero garantire che gli LMM e le applicazioni usate nella sanità e nella medicina, indipendentemente dal rischio o dal beneficio associati alle tecnologie di IA, soddisfino i principi etici e gli standard sui diritti umani che riguardano, per esempio, la dignità, l'autonomia e la privacy di una persona**
- **I governi dovrebbero emanare leggi e politiche che impongano a fornitori e sviluppatori di condurre valutazioni di impatto degli LMM e delle applicazioni, con riferimento a etica, diritti umani, sicurezza e protezione dei dati, durante tutto il ciclo di vita di un sistema di IA. Le valutazioni di impatto dovrebbero essere validate da una terza parte indipendente prima e dopo l'introduzione di una tecnologia di IA e dovrebbero essere di dominio pubblico**
- **L'onere normativo a carico di un fornitore dovrebbe aumentare se il prodotto o l'applicazione diverge sostanzialmente o modifica il modello di base in modi che sono fuori dal controllo dello sviluppatore del modello.**

- **I governi dovrebbero garantire che, per un LMM o un'applicazione regolamentata come dispositivo medico, l'onere della prova che il dispositivo funzioni come commercializzato e che soddisfi i requisiti delle leggi del Paese sia a carico dello sviluppatore e/o del fornitore**
- **I governi dovrebbero garantire che gli LMM o le applicazioni di supporto clinico alle decisioni che non sono ancora approvate per l'uso non vengano impiegate in via sperimentale al di fuori di un contesto di sperimentazione clinica autorizzata. I governi potrebbero facilitare usi sperimentali degli LMM in condizioni predefinite che consentano di valutare questi strumenti in ambienti dal vivo in contesti clinici reali, con salvaguardie e supervisione al fine di proteggere il sistema sanitario da rischi o conseguenze indesiderate**
- **I governi dovrebbero far rispettare le leggi sulla tutela dei consumatori per garantire che qualsiasi conseguenza negativa dell'uso degli LMM e delle applicazioni non si rifletta sugli utenti, inclusi i pazienti. Le leggi sulla tutela dei consumatori potrebbero essere applicate, per esempio, per prevenire pratiche equivalenti alla manipolazione di un utente o per affrontare le cause di altre conseguenze negative degli LMM o delle applicazioni al fine di proteggere i pazienti e le loro famiglie da qualsiasi danno attuale o futuro.**

6. Distribuzione di modelli di base per finalità generali

Anche quando un LMM o un'applicazione con un LMM siano stati progettati eticamente e abbiano superato un adeguato controllo normativo, possono comunque comportare rischi una volta commercializzati. Il distributore di un'applicazione o strumento di IA nel settore sanitario potrebbe essere sia lo sviluppatore sia il fornitore dell'LMM o dell'applicazione o, per esempio, un ministero della salute, un ospedale, un'azienda sanitaria o un'azienda farmaceutica.

6.1 Rischi da gestire nella distribuzione di un servizio o applicazione sanitaria con un modello di base per finalità generali

I rischi durante la distribuzione possono essere causati dall'imprevedibilità degli LMM e dalle risposte che essi generano, dalla possibilità che un modello di base per finalità generali venga utilizzato in modi non previsti né dallo sviluppatore né dal fornitore, e dal fatto che le risposte fornite da un LMM possano variare nel tempo.
I principali rischi che devono essere gestiti nell'implementazione di un LMM sono:

- **le risposte inaccurate o false**
- **i bias**
- **la privacy dei dati inseriti e generati da un LMM**
- **l'accessibilità e l'economicità di un LMM**
- **gli impatti sul lavoro e sull'occupazione**
- **i bias di automazione e la degradazione delle competenze**
- **la qualità delle interazioni tra operatori sanitari e pazienti.**

Questa sezione descrive come gli attori nella catena di valore dell'IA, inclusi gli utenti, possono mitigare o prevenire i rischi e il ruolo dei governi nella regolamentazione dell'uso degli strumenti di IA una volta che un LMM è stato rilasciato, formando i lavoratori e altri attori dei sistemi sanitari per massimizzare l'uso appropriato di un LMM.

6.2 Responsabilità continua di sviluppatori e fornitori durante la distribuzione

Sviluppatori e fornitori hanno responsabilità e obblighi anche dopo che un LMM o un'applicazione sono stati approvati per l'uso, sia perché lo sviluppatore o il fornitore distribuiscono l'LMM sia per-

ché determinati rischi possono essere gestiti dopo la distribuzione e solo da uno sviluppatore o un fornitore. Tali obblighi andrebbero definiti da regolamenti o leggi per garantire che sviluppatori e fornitori allochino risorse e attenzione adeguate a questo aspetto.

In primo luogo, i governi dovrebbero introdurre verifiche post rilascio obbligatorie e valutazioni d'impatto anche per la protezione dei dati e i diritti umani condotte da terze parti indipendenti quando un LMM viene distribuito su larga scala [155, 160]. Le verifiche post rilascio e le valutazioni d'impatto dovrebbero essere rese pubbliche e dovrebbero considerare gli esiti e gli impatti suddivisi per tipo di utente, per esempio per età, etnia o disabilità.

In secondo luogo, i governi potrebbero ritenere i fornitori e gli sviluppatori responsabili per contenuti non accurati, falsi o dannosi provenienti da un LMM dopo il suo rilascio per i quali né il fornitore né lo sviluppatore hanno preso misure per correggerli o evitarli. La regolamentazione del governo cinese sull'IA generativa, per esempio, stabilisce che un LMM non deve produrre informazioni che siano "false e dannose" [151]. Nell'Unione Europea, l'aggiunta di un LMM a un prodotto o servizio potrebbe creare ulteriori responsabilità per lo sviluppatore e il fornitore dell'LMM. Per esempio, se un LMM è integrato in un servizio che ricade nell'ambito della regolamentazione dei servizi digitali, come l'*European Union Digital Service Act*, l'LMM è indirettamente soggetto a controllo normativo, che potrebbe richiedere una supervisione regolatoria a causa della tendenza degli LMM a generare "allucinazioni" [120]. In terzo luogo, uno sviluppatore e un fornitore potrebbero essere tenuti a obblighi operativi continui nel tempo affinché governi e utenti possano utilizzare un LMM in modo sicuro. Questi potrebbero includere la fornitura di una documentazione tecnica sufficiente [133, 134].

6.3 Responsabilità dei distributori

Anche i soggetti che distribuiscono un modello sono responsabili di evitare o mitigare i rischi associati all'uso di un LMM o di un'applicazione.

In primo luogo, un distributore di un modello dovrebbe utilizzare le informazioni fornite dagli sviluppatori o dai fornitori per decidere di non usare un LMM o un'applicazione in un contesto non appropriato, per via di bias nei dati di addestramento, di bias di contesto che rendono l'LMM inadatto al contesto in cui deve essere usato o di altri errori evitabili o rischi potenziali noti al distributore. Se un distributore riceve informazioni chiare e complete su tali rischi e comunque offre l'LMM per l'uso in contesti non appropriati, dovrebbe essere ritenuto responsabile per qualsiasi danno risultante.

In secondo luogo, i distributori dovrebbero comunicare qualsiasi rischio, di cui sono ragionevolmente consapevoli, che potrebbe derivare dall'uso di un LMM e qualsiasi errore che abbia danneggiato gli utenti. Tali avvisi non dovrebbero essere in caratteri piccoli o poco visibili. In alcune circostanze particolari, un distributore potrebbe essere responsabile, anche se non imposto da una legge o un regolamento, della sospensione dell'uso o della rimozione di un LMM o di un'applicazione dal mercato per evitare ulteriori danni.

In terzo luogo, i distributori possono prendere misure per migliorare l'accessibilità e la sostenibilità

economica di un LMM. Un distributore può assicurare che i prezzi o le tariffe di abbonamento per l'uso di un LMM corrispondano alla capacità di pagamento di un governo o di un singolo utente e dovrebbe garantire che gli LMM siano addestrati e forniti in lingue e scritture che possano essere usate da persone altrimenti ignorate o escluse dai benefici della tecnologia. I distributori dovrebbero anche richiedere a fornitori e sviluppatori di garantire che gli LMM attuali e futuri siano disponibili in un ampio numero di lingue.

6.4 Programmi governativi e pratiche

L'introduzione degli LMM nei sistemi sanitari e per altri usi associati all'assistenza sanitaria richiede un significativo adeguamento da parte dei professionisti sanitari. Né gli sviluppatori né i fornitori hanno l'interesse, le risorse o l'esperienza per garantire un uso appropriato di un LMM da parte dei professionisti sanitari o per altri usi che coinvolgano persone con formazione o esperienza specializzata.

Come nella progettazione di un modello di base per finalità generali, i governi potrebbero reclutare sia i professionisti sanitari sia i pazienti in "collegi di supervisione umana" per garantire che i nuovi LMM e le applicazioni utilizzate nel processo decisionale clinico siano utilizzati in modo appropriato e non compromettano i diritti dei pazienti[1].

Governi, università (facoltà di scienze della salute) o fornitori di servizi sanitari come gli ospedali possono anche garantire che gli operatori sanitari utilizzino un LMM per fornire assistenza clinica in modo efficace e che siano adeguatamente formati al riguardo. I professionisti sanitari e i clinici dovrebbero essere formati in particolare:

1. **a comprendere come gli LMM prendono decisioni e i limiti della comprensione di come tali decisioni vengono prese**
2. **a identificare preoccupazioni sull'uso appropriato degli LMM**
3. **sui metodi per evitare il bias di automazione**
4. **a coinvolgere ed educare i pazienti che considerano o potrebbero considerare la possibilità di usare un LMM**
5. **ai rischi di cybersicurezza associati all'uso di LMM.**

La formazione e l'educazione continua dei professionisti sanitari sono di particolare importanza per informare i pazienti, i cittadini e altre terze parti quando un suggerimento è generato da un LMM o quando le informazioni fornite da un LMM sono state utilizzate per prendere una decisione medica o per un'altra attività medica. In tali comunicazioni, il paziente o la persona comune dovrebbe essere pienamente informato dei rischi associati all'uso degli LMM per preservare il suo diritto al consenso informato.

La formazione dei professionisti sanitari è fondamentale anche per garantire che, quando utilizzano un LMM a livello professionale, le loro azioni non violino inconsapevolmente le leggi, specialmente

[1] Comunicazione di David Gruson, esperto OMS sull'etica e sulla governance dell'IA per la salute

quelle relative alla protezione dei dati e delle informazioni sanitarie. Per esempio, i fornitori medici di servizi che introducono "informazioni sanitarie protette" in un chatbot con LMM potrebbero violare leggi come l'*Health Insurance Portability and Accountability Act* negli Stati Uniti [150]. Man mano che gli LMM popolari diventeranno "affidabili" per i professionisti sanitari, per esempio, potrebbero essere divulgati più dati dei pazienti di quanto ci si renda conto [154].

Gli altri portatori di interesse nel sistema sanitario dovrebbero essere istruiti sui benefici, i rischi, gli utilizzi e le sfide degli LMM nell'assistenza sanitaria e su come gli LMM si differenziano da altre tecnologie per la generazione di informazioni o consigli e come sono stati utilizzati per altri scopi nell'assistenza sanitaria. Andrebbe migliorata anche su ampia scala, la consapevolezza pubblica sull'uso degli LMM nell'assistenza sanitaria e in altri domini. La guida dell'OMS sull'etica e la governance dell'IA per la salute [1] raccomanda che: "il pubblico dovrebbe essere coinvolto nello sviluppo dell'IA per la salute per comprendere le forme di condivisione e utilizzo dei dati, per giudicare le forme di IA che sono socialmente e culturalmente accettabili e per esprimere pienamente la propria preoccupazione e le aspettative. Inoltre, l'alfabetizzazione del pubblico generale riguardo alla tecnologia di IA deve essere migliorata per consentire alle persone di scegliere quali tecnologie di IA sono accettabili".

I governi che forniscono un LMM o un'applicazione a un sistema sanitario potrebbero utilizzare la loro autorità di approvvigionamento per promuovere certe pratiche tra sviluppatori, fornitori e distributori. L'acquisizione di un LMM o di un'applicazione fondamentale per l'uso in un sistema sanitario può eliminare le barriere all'accesso e alla sostenibilità economica se la tecnologia di IA non sposta altri investimenti sanitari che potrebbero essere più efficaci, equi e accessibili. L'acquisizione pubblica può stabilire requisiti per la trasparenza rispetto all'addestramento dei dati, all'assicurazione della qualità, alla valutazione dei rischi, alla mitigazione e alle valutazioni esterne. Tali requisiti possono essere critici se un Paese non dispone né di una legislazione al riguardo né di un'agenzia regolatoria con sufficienti risorse per regolamentare efficacemente gli LMM.

Raccomandazioni:

- **Quando un LMM viene distribuito su larga scala i governi dovrebbero introdurre valutazioni post rilascio e d'impatto obbligatorie, anche per la protezione dei dati e dei diritti umani, effettuate da terze parti indipendenti. Queste valutazioni d'impatto dovrebbero essere rese pubbliche e includere esiti ed impatti suddivisi per tipo di utente, per esempio per età, etnia o disabilità**
- **I governi potrebbero ritenere i fornitori o gli sviluppatori responsabili dei contenuti imprecisi, falsi o dannosi generati da un LMM dopo il suo rilascio e che non sono stati corretti o evitati né dal fornitore né dallo sviluppatore**
- **I governi dovrebbero richiedere obblighi operativi continui nel tempo sia da parte degli sviluppatori sia dei fornitori per garantire che gli LMM e le applicazioni possano essere utilizzati in sicurezza. Questi dovrebbero includere una documentazione tecnica adeguata**
- **In conformità con le informazioni ottenute sia dagli sviluppatori sia dai fornitori, i distributori non dovrebbero utilizzare un LMM o un'applicazione in un contesto inappropriato per via di bias nei dati di addestramento e bias contestuali che rendono l'LMM inadatto in un determinato contesto, o altri errori potenziali o rischi, come contenuti non accurati, falsi o dannosi pubblicati da un LMM, che sono noti al distributore e che possono essere evitati**
- **I distributori dovrebbero comunicare qualsiasi rischio di cui ragionevolmente sono a conoscenza che potrebbe derivare dall'uso di un LMM, così come errori che hanno causato danni agli utenti; tali avvertimenti non dovrebbero essere in caratteri piccoli o poco visibili. In alcune circostanze, un distributore può essere ritenuto responsabile, anche se non richiesto da una legge o regolamento, della sospensione o del ritiro dal mercato di un l'LMM o di una applicazione per evitare danni futuri**
- **I distributori dovrebbero migliorare l'accessibilità e la sostenibilità economica di un LMM, assicurando che i prezzi o le tariffe di abbonamento per l'uso siano in linea con le capacità economiche di un governo o di un singolo utente e dovrebbero garantire che gli LMM siano addestrati e offerti in lingue e scritture che possano essere utilizzate da persone altrimenti ignorate o escluse dai benefici della tecnologia. I distributori dovrebbero, inoltre, richiedere ai fornitori e agli sviluppatori di garantire che gli LMM attuali e futuri siano sviluppati in varie lingue**
- **I governi dovrebbero facilitare la partecipazione di professionisti sanitari e pazienti, in "collegi di supervisione umana" per garantire che i nuovi LMM e le applicazioni utilizzate nel prendere decisioni cliniche siano utilizzati in modo appropriato e non compromettano i diritti dei pazienti**
- **I ministeri della salute e le università (facoltà di scienze della salute) dovrebbero formare i professionisti sanitari e i clinici: a comprendere come gli LMM prendono decisioni e i limiti della comprensione di come tali decisioni vengono prese), a identificare e comprendere le preoccupazioni su un loro uso appropriato, sui metodi per evitare il bias di automazione, a coinvolgere ed educare i pazienti che stanno considerando l'uso degli LMM e sui rischi di cybersicurezza associati all'uso degli LMM**
- **I governi, i fornitori di servizi sanitari, i ricercatori sanitari e i finanziatori dovrebbero coinvolgere i cittadini affinché partecipino a diverse forme di condivisione e utilizzo dei dati, possano commentare se e come gli LMM siano socialmente e culturalmente accettabili ed esprimere pienamente le proprie preoccupazioni e aspettative. Inoltre, dovrebbe essere migliorata l'alfabetizzazione del pubblico nella tecnologia di IA per consentire di identificare gli usi e i tipi accettabili di LMM**
- **I governi che forniscono un LMM o un'applicazione nel sistema sanitario dovrebbero garantire che la loro autorità di approvvigionamento promuova certe garanzie da parte di sviluppatori, fornitori e distributori, inclusa la trasparenza.**

7. Responsabilità per gli LMM

Con l'ampliarsi dell'uso degli LMM nell'assistenza sanitaria e nella medicina, errori, usi impropri e, infine, danni alle persone sono inevitabili. Sarà necessario ricorrere a regolamentazioni sulla responsabilità per compensare le persone che hanno subìto danni, stabilendo nuove forme di correzione quando gli approcci attuali sono insufficienti o obsoleti.
La progettazione, lo sviluppo, la valutazione di qualità e la distribuzione delle tecnologie di IA coinvolgono varie entità, ognuna delle quali svolge un ruolo distinto. Questo può complicare l'attribuzione della responsabilità. Gli sviluppatori possono esigere che le entità a valle, come fornitori e distributori, siano responsabili per qualsiasi danno risultante dall'uso di un LMM, mentre le entità a valle possono sostenere che azioni precedenti, come la scelta dei dati utilizzati per addestrare un algoritmo, siano la causa.

Sviluppatori e fornitori possono anche affermare che, una volta che una tecnologia di IA medica sia stata approvata per l'uso da un ente regolatore, non dovrebbero più essere ritenuti responsabili per eventuali danni causati (normativa preventiva) [1]. Stabilire la responsabilità lungo la catena di valore è una sfida per legislatori e responsabili delle politiche.

Le regole per la responsabilità civile devono garantire che una vittima di danni possa richiedere compensazione e risarcimento, per quanto possa essere difficile assegnare colpa e responsabilità tra le entità coinvolte nello sviluppo e nella distribuzione di una tecnologia di IA. Se le vittime hanno di fronte un percorso troppo difficile per ottenere una compensazione, non può esserci giustizia e nessun incentivo per le parti nella catena di valore per evitare tali danni in futuro. Le regole dovrebbero anche garantire che la compensazione sia adeguata al danno subìto.

L'Unione Europea nella sua proposta di Direttiva sulla responsabilità dell'IA (*AI Liability Directive*) semplifica l'onere della prova per una vittima introducendo una "presunzione di causalità" [162]. Così, se una vittima può dimostrare che una o più entità non hanno rispettato un obbligo rilevante per il danno e che è probabile un collegamento causale con la prestazione dell'IA, il tribunale può presumere che la non conformità sia stata la causa del danno [162]. L'onere è quindi posto in capo alla parte responsabile per confutare la presunzione, per esempio indicando un'altra parte come causa del danno. L'ambito della legislazione non è limitato al produttore originale di un sistema di IA ma include qualsiasi partecipante nella catena di valore dell'IA [162]. Quando tutti gli attori nella catena del valore sono considerati congiuntamente responsabili, possono dimostrare la loro efficienza nel gestire e mitigare i rischi al fine di ridurre la propria responsabilità.
Tuttavia, è possibile che non si riesca a chiarire di chi sia la responsabilità e chi deve provvedere al risarcimento per le lesioni causate da prodotti e servizi guidati dall'IA, specialmente se la persona

non sa che è stato usato un LMM per prendere una decisione clinica nel suo caso. Le nuove regole possono non essere complete per quanto riguarda la responsabilità per le lesioni causate dalle tecnologie mediche guidate dall'IA [163].

Poiché gli LMM sono altamente speculativi, poco compresi e guidati dalle pressioni economiche del mercato, i governi potrebbero considerare gli LMM utilizzati nell'assistenza sanitaria come prodotti per i quali sviluppatori, fornitori e distributori saranno tenuti a uno standard di responsabilità rigoroso. Rendere responsabili questi attori per qualsiasi errore potrebbe garantire che un paziente venga risarcito se l'errore li riguarda [1], ma ciò dipende dal fatto che il paziente sia informato che nel suo caso è stato utilizzato un LMM. Sebbene questa continua responsabilità possa scoraggiare l'uso di LMM sempre più sofisticati, potrebbe anche moderare la volontà di assumere rischi non necessari e di rilasciare nuovi LMM in contesti sanitari o di salute pubblica prima che i loro molti rischi e potenziali danni siano stati pienamente identificati e affrontati [1].

Il sistema di responsabilità per l'IA potrebbe tuttavia non essere adeguato per attribuire la colpa, perché gli algoritmi si evolvono in modi che né gli sviluppatori, né i fornitori né i distributori possono controllare completamente. Inoltre, potrebbero esserci situazioni o giurisdizioni in cui una persona danneggiata non ha la possibilità di recuperare i danni. Per esempio, negli Stati Uniti un paziente che subisce un danno utilizzando direttamente un LMM per cercare un parere medico potrebbe non avere modo di ottenere un risarcimento perché i sistemi di IA non sono inclusi nelle regole di responsabilità professionale, ed eccezioni o limitazioni alle leggi sulla responsabilità del prodotto o del consumatore possono precludere il riconoscimento dei danni subiti [163]. In altri ambiti dell'assistenza sanitaria, talvolta viene fornita una compensazione senza l'attribuzione di colpa o responsabilità, come per i danni medici risultanti da effetti avversi dei vaccini. La guida dell'OMS raccomandava di determinare "se i fondi di compensazione senza colpa, senza responsabilità, sono un meccanismo appropriato per fornire pagamenti alle persone che subiscono danni medici a causa dell'uso delle tecnologie di IA, incluse le modalità per rendere disponibili risorse per far fronte a eventuali richieste" [1]. Tale raccomandazione è valida anche oggi e potrebbe essere un mezzo per determinare il compenso per danni causati dagli LMM o dalle applicazioni con LMM.

Raccomandazione:

- **I governi dovrebbero stabilire la responsabilità lungo la catena di valore dello sviluppo, della fornitura e della distribuzione degli LMM e delle applicazioni per garantire che la vittima di un danno dovuto a un LLM possa richiedere una compensazione, indipendentemente dalla difficoltà di attribuire la colpa e dalle responsabilità delle diverse entità coinvolte nello sviluppo e nella distribuzione della tecnologia.**

8. Governance internazionale degli LMM

I governi dovrebbero sostenere lo sviluppo collettivo di regole internazionali per la governance degli LMM e di altre forme di IA utilizzate nell'assistenza sanitaria, poiché tali tecnologie stanno proliferando a livello globale. Un esempio è rappresentato dalla strategia globale dell'OMS sulla salute digitale 2020-2025. Il processo dovrebbe coinvolgere una maggiore cooperazione e collaborazione all'interno del sistema delle Nazioni Unite per rispondere alle opportunità e alle sfide della distribuzione dell'IA nell'assistenza sanitaria e della sua diffusione su vasta scala nella società e nell'economia. A meno che i governi non collaborino per stabilire standard appropriati e applicabili, il numero di LMM e di altre forme di IA che non rispettano gli standard legali, etici e di sicurezza potrebbe aumentare, causando possibili danni se non saranno introdotte o applicate adeguatamente normative e altre forme di protezione, sia volontariamente sia a causa di risorse insufficienti. L'OMS ha recentemente pubblicato una nuova guida, in cooperazione con le agenzie regolatorie di tutto il mondo, che illustra i principi fondamentali che i governi e le autorità regolatorie possono seguire per sviluppare nuove linee guida o adattare quelle esistenti sull'IA [164].

La governance internazionale può scongiurare una "corsa al ribasso" tra le aziende alla ricerca di un vantaggio da pioniere, dove vengono ignorati gli standard di sicurezza ed efficacia, e tra i governi, in cerca di vantaggi nella competizione geopolitica per la supremazia tecnologica. Pertanto, la governance internazionale può garantire il rispetto degli standard minimi di sicurezza ed efficacia da parte di tutte le aziende e può impedire l'introduzione di regolamentazioni che conferiscono vantaggi competitivi o svantaggi ad aziende o governi. La governance internazionale può rendere i governi responsabili dei loro investimenti e della loro partecipazione allo sviluppo e alla distribuzione di sistemi basati sull'IA, garantendo l'introduzione di norme appropriate che rispettino i principi etici, i diritti umani e il diritto internazionale. L'assenza di standard applicabili a livello internazionale potrebbe influenzare negativamente l'adozione di tali tecnologie.

La governance internazionale potrebbe assumere varie forme. Una proposta è quella di istituire un'agenzia di ricerca pubblica, finanziata dai vari governi, analoga all'Organizzazione europea per la ricerca nucleare (CERN), una collaborazione internazionale, con finanziamenti e risorse per perseguire progetti di trasformazione di ampia portata, i cui risultati sono condivisi pubblicamente [165, 166]. In alternativa, è stato suggerito che tale entità venga incaricata di sviluppare le forme più avanzate e rischiose di IA in un ambiente altamente sicuro, rendendo illegittimi altri tentativi di costruire queste forme di IA [167]. Attualmente, tali progetti su larga scala non rientrano nel dominio dei progetti finanziati dalle pubbliche istituzioni per generare beni comuni, ma sono di competenza delle grandi aziende tecnologiche in concorrenza commerciale tra loro. Alcuni leader, inclusi leader

mondiali e dirigenti di aziende tecnologiche, hanno sottolineato la necessità di trattare l'IA in modo simile alle armi nucleari, adottando un quadro normativo globale simile ai trattati per l'uso di queste armi [109].

Indipendentemente dalla forma di governance internazionale adottata, è imperativo che non sia plasmata esclusivamente da Paesi ad alto reddito o da Paesi che lavorano principalmente o esclusivamente con le più grandi aziende tecnologiche del mondo [168]. Gli standard sviluppati da e per i Paesi ad alto reddito e le aziende tecnologiche, sia per tutte le applicazioni di IA sia per l'uso specifico degli LMM nell'assistenza sanitaria e nella medicina, potrebbero escludere dalla definizione degli standard la maggior parte dell'umanità nei Paesi a basso e a medio reddito. Ciò potrebbe rendere le future tecnologie dell'IA potenzialmente pericolose o inefficaci nei Paesi che potrebbero trarne maggior beneficio.

La governance internazionale dell'IA richiede la cooperazione di tutti gli attori attraverso un multilateralismo in rete, come proposto nel 2019 dal Segretario generale delle Nazioni Unite [169], che coinvolgerebbe le Nazioni Unite, istituti finanziari internazionali, organizzazioni regionali, blocchi commerciali e altri soggetti, inclusa la società civile, le città, le imprese, le autorità locali e i giovani, al fine di lavorare più strettamente, efficacemente e inclusivamente. Porre l'etica e i diritti umani al centro dello sviluppo e della distribuzione degli LMM può contribuire in modo sostanziale al raggiungimento della copertura sanitaria universale.

Raccomandazione:

- **I governi dovrebbero sostenere lo sviluppo collettivo di regole internazionali per la governance dell'IA. Qualunque sia la forma di governance, non dovrebbe essere plasmata esclusivamente da Paesi ad alto reddito o da Paesi che lavorano principalmente o esclusivamente con le più grandi aziende tecnologiche del mondo, poiché tale approccio lascerebbe la maggior parte dell'umanità, nei Paesi a basso e a medio reddito, senza un ruolo o una voce nella definizione della governance internazionale dell'IA.**

Riferimenti bibliografici

1. Ethics and governance of artificial intelligence for health. Geneva, World Health Organization, 2021 (https://www.who.int/publications/i/item/9789240029200, accessed 26 May 2023).

2. Khullar D. Can AI treat mental illness? The New Yorker, 27 February 2023 (https://www.newyorker.com/magazine/2023/03/06/can-ai-treat-mental-illness, accessed 29 May 2023).

3. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med 2022;28(9):1773-84.

4. Hariri Y, Harris T, Raskin A. You can have the blue pill or the red pill, and we're out of blue pills. The New York Times, 24 March 2023 (https://www.nytimes.com/2023/03/24/opinion/yuval-harari-ai-chatgpt.html, accessed 26 May 2023).

5. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ et al. Foundation models for generalist medical artificial intelligence. Nature 2023;616:259-65.

6. Hu K. ChatGPT sets record for fastest growing user-base - analyst note. Reuters, 2 February 2023 (https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/, accessed 26 May 2023).

7. Weise K, Grant N. Microsoft and Google unveil A.I. tools for businesses. The New York Times, 16 March 2023 (https://www.nytimes.com/2023/03/16/technology/microsoft-google-ai-tools-businesses.html, accessed 26 May 2023).

8. Yang Z. Chinese tech giant Baidu just released its answer to ChatGPT. MIT Technology Review, 16 March 2023 (https://www.technologyreview.com/2023/03/16/1069919/baidu-ernie-bot-chatgpt-launch/, accessed 26 May 2023).

9. Murgia M, Bradshaw T. Musk to launch AI start-up to rival ChatGPT. Financial Times, 15 April 2023 (https://www.ft.com/content/2a96995b-c799-4281-8b60-b235e84aefe4, accessed 26 May 2023).

10. Heaven WD. The open source AI boom is built on Big Tech's handouts. How long will it last? MIT Technology Review, 12 May 2023 (https://www.technologyreview.com/2023/05/12/1072950/open%20source-ai-google-openai-eleuther-meta, accessed 26 May 2023).

11. Martin A. Google CEO Sunder Pichai admits people don't fully understand how chatbot AI works, Evening Standard, 17 April 2023 (https://www.standard.co.uk/tech/google-ceo-sundar-pichai-understand-ai-chatbot-bard-b1074589.html, accessed 26 May 2023).

12. Roose K. A conversation with Bing's chatbot left me deeply unsettled. The New York Times, 16 February 2023 (https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html, accessed 26 May 2023).

13. Marcus G. AI platforms like ChatGPT are easy to use but potentially dangerous, Scientific American, 19 December 2022 (https://www.scientificamerican.com/article/ai-platforms-like-chatgpt-are-easy-to-use-but-also-potentially-dangerous/, accessed 26 May 2023).

14. Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E et al. Sparks of artificial general intelligence: early experiments with GPT-4. ArXiv:2302.12712.

15. McGowran L. OpenAI criticised for lack of transparency around ChatGPT-4. Silicon Republic, 16 March 2023 (https://www.siliconrepublic.com/machines/openai-gpt4-transparency-ai-concerns-stripe-chatgpt, accessed 26 May 2023).

16. Spitale G, Biller-Andorno N, Germani F. AI model GPT-3 (dis)informs us better than humans. Sci Adv 2023;DOI:10.1126/sciadv.adh1850.

17. Volpicelli G. ChatGPT broke the EU plan to regulate AI. Politico, 3 March 2023 (https://www.politico.eu/article/eu-plan-regulate-chatgpt-openai-artificial-intelligence-act/, accessed 26 May 2023).

18. Arcesati R, Chang W. China is blazing a trail in regulating Generative AI - on the CCP'sterms. The Diplomat, 28 April 2023 (https://thediplomat.com/2023/04/china-is-blazing-a-trail-in-regulating-generative-ai-on-the-ccps-terms/, accessed 26 May 2023).

19. Martindale J. These are the countries where ChatGPT is currently banned. Digital Trends, 12 April 2023 (https://www.digitaltrends.com/computing/these-countries-chatgpt-banned/, accessed 26 May 2023).

20. Johnson K. ChatGPT can help doctors – and hurt patients. Wired, 24 April 2023 (https://www.wired.com/story/chatgpt-can-help-doctors-and-hurt-patients/, accessed 28 May 2023).

21. Topol E. Multimodal AI for medicine, simplified. Ground Truths, 14 March 2023 (https://erictopol.substack.com/p/multimodal-ai-for-medicine-simplified, accessed 28 May 2023).

22. Heaven WD. AI hype is built on high test scores. Those tests are flawed. MIT Technology Review, 30 August 2023 (https://www.technologyreview.com/2023/08/30/1078670/large-language-models-arent-people-lets-stop-testing-them-like-they-were/, accessed 1 October 2023).

23. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW et al. Large language models encode clinical knowledge. Nature 2023;620:172-80.

24. Kulkarni PA, Singh H. Artificial intelligence in clinical diagnosis: opportunities, challenges, and hype. JAMA 2023;330:317-8.

25. Subbamaran N. ChatGPT will see you now: doctors using AI to answer patient questions. Wall Street Journal, 28 April 2023 (https://www.wsj.com/articles/dr-chatgpt-physicians-are-sending-patients-advice-using-ai-945cf60b, accessed 28 May 2023).

26. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023;183:589-96.

27. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. New Engl J Med 2023;388:1233-9.

28. The potential of large language models in healthcare: improving quality of care and patient outcomes, Medium, 7 December 2022. (https://medium.com/@BuildGP/the-potential-of-large-language-models-in-healthcare-improving-quality-of-care-and-patient-6e8b6262d5ca, accessed 28 May 2023).

29. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T et al. CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv. 2017;DOI:10.48550/arXiv.1711.05225.

30. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C et al. A large language model for electronic health records. NPJ Digit Med 2022;DOI:10.1038/s41746-022-00742-2.

31. Ghahramani Z. Introducing PaLM 2. The Keyword, 10 May 2023 (https://blog.google/technology/ai/google-palm-2-ai-large-language-model/, accessed 28 May 2023).

32. Weise K, Metz C. When A.I. chatbots hallucinate. The New York Times, 9 May 2023 (https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html, accessed 1 June 2023).

33. Bender EM, Gebru T, McMillan-Major A, Mitchell M. On the dangers of stochastic parrots: can language models be too big? In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, March 2021;DOI:10.1145/3442188.3445922.

34. Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. Lancet Digit Health 2023;DOI:10.1016/s2589-7500 (23)00083-3.

35. Metz, Cade. Chatbots may 'hallucinate' more often than many realize. The New York Times, 6 November 2023 (https://www.nytimes.com/2023/11/06/technology/chatbots-hallucination-rates.html, accessed 7 November 2023).

36. Acar OA. AI prompt engineering isn't the future. Harvard Business Review, 6 June 2023 (https://hbr.org/2023/06/ai-prompt-engineering-isnt-the-future?registration=success, accessed 26 June 2023).

37. GPT-4 system card. Open AI, 23 March 2023 (https://cdn.openai.com/papers/gpt-4-system-card.pdf, accessed 28 May 2023).

38. GPT-4. OpenAI, 14 March 2023 (https://openai.com/research/gpt-4, accessed 28 May 2023).

39. Radford A, Kleinman Z. ChatGPT can now access up-to-date information. BBC News, 27 September 2022 (https://www.bbc.com/news/technology-66940771, accessed 1 October 2023).

40. Kruge S, Ostermaier A, Uhl M. The moral authority of ChatGPT. ArXiv:23101.07098.

41. Mickle T, Metz C, Grant N. The chatbots are here, and the internet industry is in a tizzy. The New York Times, 8 March 2023 (https://www.nytimes.com/2023/03/08/technology/chatbots-disrupt-internet-industry.html, accessed 29 May 2023).

42. Woo M. Trial by artificial intelligence. Nature 2019;573:S100-2 (https://media.nature.com/original/magazine-assets/d41586-019-02871-3/d41586-019-02871-3.pdf, accessed 29 May 2023).

43. Muralidharan V, Burgart A, Daneshjou D, Rose S. Recommendations for the use of pediatric data in artificial intelligence and machine learning ACCEPT-AI. NPJ Dig Med 2023;DOI:10.1038/s41746-023-00898-5.

44. Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learning Individual Differences 2023;DOI:10.1016/j.lindif.2023.102274.

45. Reddy CD, Lopez L, Ouyang D, Zou JY, He B. Video-based deep learning for automate assessment of left ventricular ejection fraction in pediatric patients. J Am Soc Echocardiogr 2023;36:482-9.

46. Knight W. These ChatGPT rivals are designed to play with your emotions. Wired, 4 May 2023 (https://www.wired.com/story/fast-forward-chatgpt-rivals-emotions/#:~:text=12%3A00%20PM-,These%20ChatGPT%20Rivals%20Are%20Designed%20to%20Play%20With%20Your%20Emotions,%2C%20-companionship%E2%80%94and%20even%20romance, accessed 29 May 2023).

47. Smuha NA, De Ketalaere M, Coeckelbergh M, Dewitte P, Poullet Y. Open letter: We are not ready for manipulative AI - urgent need for action. KU Leuven, 31 March 2023 (https://www.law.kuleuven.be/ai-summer-school/open-brief/open-letter-manipulative-ai, accessed 29 May 2023).

48. Cuthbertson A. "No, I'm not a robot": ChatGPT successor tricks worker into thinking it is human. Independent, 15 March 2023 (https://www.independent.co.uk/tech/chatgpt-gpt4-ai-openai-b2301523.html, accessed 26 June 2023).

49. Walker L. Belgian man dies by suicide following exchanges with chatbot. The Brussels Times, 28 March 2023 (https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt, accessed 29 May 2023).

50. DeGuerin M. Oops: Samsung employees leaked confidential data to ChatGPT. Gizmodo, 6 April 2023 (https://gizmodo.com/chatgpt-ai-samsung-employees-leak-data-1850307376, accessed 29 May 2023).

51. Privacy policy. OpenAI, 27 April 2023 (https://openai.com/policies/privacy-policy, accessed 29 March 2023).

52. Coles C. 11% of data employees paste into ChatGPT is confidential. Cyberhaven, 19 April 2023 (https://www.cyberhaven.com/blog/4-2-of-workers-have-pasted-company-data-into-chatgpt/, accessed 29 May 2023).

53. Mihalcik C. ChatGPT bug exposed some subscribers'payment info. CNET, 24 March 2023. (https://www.cnet.com/tech/services-and-software/chatgpt-bug-exposed-some-subscribers-payment-info/, accessed 29 May 2023).

54. Moodley K, Rennie S. ChatGPT has many uses. Experts explore what this means for healthcare and medical research. The Conversation, 22 February 2023 (https://theconversation.com/chatgpt-has-many-uses-experts-explore-what-this-means-for-healthcare-and-medical-research-200283, accessed 2 June 2023).

55. De Proost M, Pozzi G. Conversational artificial intelligence and the potential for epistemic injustice. Am J Bioethics 2023;23:51-3.

56. Disability and employment. New York: United Nations, Department of Economic and Social Affairs (Disability); un-

dates (https://www.un.org/development/desa/disabilities/resources/factsheet-on-persons-with-disabilities/disability-and-employment.html, accessed 11 September 2023).

57. Whittaker M, Alper M, Bennett CL, Hendren S, Kaziunas L, Mills Met al. Disability, bias and AI. New York, AI Now Institute; 2019 (https://ainowinstitute.org/wp-content/uploads/2023/04/disabilitybiasai-2019.pdf, accessed 11 September 2023).67

58. Hallman J. AI language models show bias against people with disabilities, study finds. University Park (PA), Penn State University; 2022 (https://www.psu.edu/news/information-sciences-and-technology/story/ai-language-models-show-bias-against-people-disabilities/, accessed 11 September 2023).

59. Harrer S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. eBioMedicine 2023;DOI:10.1016/j.ebiom.2023.104512.

60. Lohr S. AI may someday work medical miracles. For now, it helps do paperwork. The New York Times, 26 June 2023 (https://www.nytimes.com/2023/06/26/technology/ai-health-care-documentation.html, accessed 10 July 2023).

61. Eddy N. Epic, Microsoft partner to use generative AI for better EHRs. Healthcare IT News, 18 April 2023 (https://www.healthcareitnews.com/news/epic-microsoft-partner-use-generative-ai-better-ehrs, accessed 31 May 2023).

62. Nuance and Microsoft announce the first fully AI-automated clinical documentation application for healthcare. Burlington (MA), Nuance; 2023 (https://news.nuance.com/2023-03-20-Nuance-and-Microsoft-Announce-the-First-Fully-AI-Automated-Clinical-Documentation-Application-for-Healthcare, accessed 31 May 2023).

63. Ahn S. The impending impacts of large language models on medical education. Korean J Med Educ 2023;35:103-7.

64. Luo R, Sun L, Xia Y, Qin T, Zhang S, Poon H et al. BioGPT: generative pre-trained transformer for biomedical text generation and mining. Briefi Bioinformatics 2022;DOI:10.1093/bib/bbac409.

65. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. Drug Discov Today 2021;26:80-93.

66. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. Nature 2023;DOI:10.1038/d41586-023-00191-1.

67. Zielinksi C, Winker MA, Aggarwal R, Ferris LA, Heinemann M, Lapeña JF Jr et al. Chatbots, generative AI, and scholarly manuscripts. Overijssel: World Association of Medical Editors, 2023. (https://wame.org/page3.php?id=106, accessed 26 June 2023).

68. Gibbs W. Lost science in the Third World. Sci Am 1995;273:92-9.

69. Birhane A, Kasirzadeh A, Leslie D, Wachter S. Science in the age of large language models. Nat Rev Phys 2023;5:277-80.

70. Monitoring the building blocks of health systems: a handbook of indicators and their measurement strategies. Geneva, World Health Organization, 2010 (https://apps.who.int/iris/bitstream/handle/10665/258734/9789241564052-eng.pdf, accessed 26 June 2023).

71. Morozov E. The true threat of artificial intelligence. The New York Times, 30 June 2023 (https://www.nytimes.com/2023/06/30/opinion/artificial-intelligence-danger.html, accessed 2 July 2023).

72. Introducing ChatGPTPlus. San Francisco (CA), Open AI, 2023 (https://openai.com/blog/chatgpt-plus, accessed 1 June 2023).

73. The hidden workforce that helped filter violence and abuse out of ChatGPT. Wall Street Journal, 11 July 2023 (https://www.wsj.com/podcasts/the-journal/the-hidden-workforce-that-helped-filter-violence-and-abuse-out-of-chatgpt/ffc2427f-bdd8-47b7-9a4b-27e7267cf413, accessed 13 July 2023).

74. Firth N. Language models may be able to self-correct biases - if you ask them. MIT Technology Review, 20 March 2023 (https://www.technologyreview.com/2023/03/20/1070067/language-models-may-be-able-to-self-correct-biases-if-you-ask-them-to/, accessed 1 June 2023).

75. Khan L. We must regulate A.I. Here's how. The New York Times, 3 May 2023 (https://www.nytimes.com/2023/05/03/opinion/ai-lina-khan-ftc-technology.html, accessed 2 June 2023).

76. Hatzius J, Briggs J, Kodnani D, Pierdomenico G. The potentially large effects of artificial intelligence on economic growth (Briggs/Kodnani). Goldman Sachs Economics Research, 26 May 2023 (https://www.key4biz.it/wp-content/uploads/2023/03/Global-Economics-Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf, accessed 1 June 2023).

77. Milmo D. AI revolution puts skilled jobs at highest risk, says OECD. The Guardian, 11 July 2023 (https://www.theguardian.com/technology/2023/jul/11/ai-revolution-puts-skilled-jobs-at-highest-risk-oecd-says, accessed 12 July 2023).

78. Health and care workforce in Europe: time to act. Geneva, World Health Organization; 2022 (https://iris.who.int/handle/10665/362379, accessed 1 June 2023).

79. Health workforce. Geneva, World Health Organization; 2023 (https://www.who.int/health-topics/health-workforce#tab=tab_1, accessed 1 June 2023).

80. Hurst L. OpenAI says 80% of workers could see their jobs impacted by AI. These are the jobs most impacted, Euronews.next, 30 March 2023. (https://www.euronews.com/next/2023/03/23/openai-says-80-of-workers-could-see-their-jobs-impacted-by-ai-these-are-the-jobs-most-affe, accessed 1 June 2023).

81. A new era of generative AI for everyone. Dublin, Accenture, 2023 (https://www.accenture.com/content/dam/accenture/final/accenture-com/document/Accenture-A-New-Era-of-Generative-AI-for-Everyone.pdf, accessed 1 June 2023).

82. Burgess M. The security hole at the heart of ChatGPT and Bing. Wired, 25 May 2023 (https://www.wired.co.uk/article/chatgpt-prompt-injection-attack-security, accessed 1 June 2023).

83. Heikkila M. Open AI's hunger for data is coming back to bite it. MIT Technology Review, 19 April 2023 (https://www.technologyreview.com/2023/04/19/1071789/openais-hunger-for-data-is-coming-back-to-bite-it/, accessed 1 June 2023).

84. General Data Protection Regulation, Regulation 2016//679 of the European Parliament and of the Council, 27 April 2016. Strasbourg, European Parliament, 2016 (https://eur-lex.europa.eu/eli/reg/2016/679/oj, accessed 27 September 2023).

85. The impact of the General Data Protection Regulation on artificial intelligence (STOA Options Brief). Strasbourg, European Parliament, 2020 (https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf, accessed 26 June 2023).

86. OPC launches investigation into ChatGPT. Ottawa, Office of the Privacy Commissioner of Canada, 4 April 2023 (https://www.priv.gc.ca/en/opc-news/news-and-announcements/2023/an_230404/, accessed 1 June 2023).

87. Lomas N. Italy orders ChatGPT blocked citing data protection concerns. Tech Crunch, 31 March 2023 (https://techcrunch.com/2023/03/31chatgpt-blocked-italy/, accessed 1 June 2023).

88. ChatGPT: Italian SA to lift temporary limitation if OpenAI implements measures. Rome: Italian Data Protection Authority; 2023 (https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9874751#english, accessed 1 June 2023).

89. Weatherbed J. OpenAI's regulatory troubles are only just beginning. The Verge, 5 May 2023 (https://www.theverge.com/2023/5/5/23709833/openai-chatgpt-gdpr-ai-regulation-europe-eu-italy, accessed 1 June 2023).

90. Wiggers K. Open AI's new tool attempts to explain language models'behaviours. Tech Crunch, 9 May 2023 (https://techcrunch.com/2023/05/09/openais-new-tool-attempts-to-explain-language-models-behaviors/, accessed 1 June 2023).

91. Libeau D. ChatGPT will probably never comply with GDPR. 10 April 2023. (https://blog.davidlibeau.fr/chatgpt-will-probably-never-comply-with-gdpr/, accessed 1 June 2023).

92. Lomas N. ChatGPT maker OpenAI accused of string of data protection breaches in GDPR complaint filed by privacy researcher. TechCrunch, 30 August 2023. (https://consent.yahoo.com/v2/collectConsent?sessionId=3_cc-session_6bdecae4-d7b6-448f-8e26-e7805c03b964, accessed 11 September 2023).

93. Fung B. The FTC should investigate Open AI and block GPT over “deceptive” behaviour, AI policy group claims. CNN, 30 March 2023. (https://edition.cnn.com/2023/03/30/tech/ftc-openai-gpt-ai-think-tank/index.html, accessed 2 June 2023).

94. Waters R, Murgia M, Espinoza J. Open AI warns over split with Europe as AI regulation advances. Financial Times, 25 May 2023 (https://www.ft.com/content/5814b408-8111-49a9-8885-8a8434022352, accessed 1 June 2023).

95. Technology-facilitated gender-based violence: making all spaces safe. New York: United Nations Population Fund, 2021 (https://www.unfpa.org/publications/technology-facilitated-gender-based-violence-making-all-spaces-safe, accessed 1 October 2023).

96. Murgia M. DeepMind reinvents itself for AI counterattack. Financial Times, 2 May 2023 (https://ft.pressreader.com/v99c/20230502/281724093873699, accessed 2 June 2023).

97. Schaake M. Regulating AI will put companies and governments at loggerheads, Financial Times, 2 May 2023 (https://www.ft.com/content/7ef4811d-79bb-4b4f-b28f-b46430f0c9ff, accessed 2 June 2023).

98. Metz, Cade. Tech giants are paying huge salaries for scarce A.I. talent. The New York Times, 22 October 2017 (https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-salaries.html).

99. Leswing K. Google reveals its newest AI supercomputer, says it beats Nvidia. CNBC, 5 April 2023. (https://www.cnbc.com/2023/04/05/google-reveals-its-newest-ai-supercomputer-claims-it-beats-nvidia-.html, accessed 2 June 2023)

100. Ahuja K. Antitrust has role in policing AI landscape. Financial Times, 10 April 2023 (https://www.ft.com/content/953817f5-5bc4-49e1-b583-977cc4780eca, accessed 2 June 2023).

101. Ahmed N, Wahed M, Thompson NC. The growing influence of industry in AI research. Science. 2023;379:884-6.

102. Røttingen JA, Regmi S, Eide M, Young AJ, Viergever RF, Ardal C et al. Mapping of available health research and development data: what’s there, what’s missing, and what role is there for a global observatory? Lancet 2013;382:1286-307.

103. A new partnership to promote responsible AI. Google Blogs, 26 July 2023 (https://blog.google/outreach-initiatives/public-policy/google-microsoft-openai-anthropic-frontier-model-forum/#:~:text=Anthropic%2C%20Google%2C%20Microsoft%20and%20OpenAI%20are%20launching%20the%20Frontier%20Model,development%20of%20frontier%20AI%20models, accessed 29 July 2023).

104. Fact sheet: Biden-Harris Administration secures voluntary commitments from leading artificial intelligence companies to manage the risks posed by AI. Washington DC, The White House, 21 July 2023 (https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/, accessed 29 July 2023).

105. Volpicelli G. Europe pitches AI pact to curtail the booming tech’s risk. Politico, 26 May 2023 (https://www.politico.eu/article/big-tech-rumble-europe-global-artificial-intelligence-debate-ai-pact/, accessed 29 July 2023).

106. Grant N, Weise K. In AI race, Microsoft and Google choose speed over caution. The New York Times, 7 April 2023 (https://www.nytimes.com/2023/04/07/technology/ai-chatbots-google-microsoft.html, accessed 2 June 2023).

107. Center for Research on Foundation Models. The Foundation Model Transparency Index, 2023. (https://crfm.stanford.edu/fmti/, accessed 21 October 2023).

108. Schiffer Z, Newton C. Microsoft lays off team that taught employees how to make AI tools responsibly. The Verge, 14 March 2023. (https://www.theverge.com/2023/3/13/23638823/microsoft-ethics-society-team-responsible-ai-layoffs, accessed 2 June 2023).

109. Milmo D. Google chief warns AI could be harmful if deployed wrongly. The Guardian, 17 April 2023 (https://www.theguardian.com/technology/2023/apr/17/google-chief-ai-harmful-sundar-pichai, accessed 2 June 2023).

110. Fiesler C. AI has social consequences, but who pays the price? Tech companies’problem with ethical debt. The Conversation, 19 April 2023 (https://theconversation.com/ai-has-social-consequences-but-who-pays-the-price-tech-companies-problem-with-ethical-debt-203375, accessed 2 June 2023).

111. Criddle, Cristina and Murphy, Hannah, Meta disbands protein-folding team in shift towards commercial AI, Financial Times, 7 August 2023. (https://www.ft.com/content/919c05d2-b894-4812-aa1a-dd2ab6de794a?accessToken=z-

wAGBZu-oWVwkdORnAXSuJRIEtOqGt0qtt55Sg.MEQCIA1QQ1iG8KPAnuDAuPvt-Ngds3OzxL1lt-0FnaVbAQFtAiAZvHnmKD_fABj8ZzLTNXRp1v7V38nTcUf_pPxAPdx16A&sharetype=gift&token=3ac5a132-e08e-412e-bc3c-08edea8a7417, accessed 18 September 2023).

112. Ananthaswamy A. In AI, is bigger always better? Nature, 8 March 2023 (https://www.nature.com/articles/d41586-023-00641-w, accessed 2 June 2023).

113. Li P. Making AI less "thirsty": uncovering and addressing the secret water footprint of AI models. ArXiv 2023;2304.03271v.

114. Syed N. The secret water footprint of AI technology. The Markup, 15 April 2023 (https://themarkup.org/hello-world/2023/04/15/the-secret-water-footprint-of-ai-technology, accessed 2 June 2023).

115. Livingstone G. It's pillage: thirsty Uruguayans blast Google's plan to exploit water supply. The Guardian, 11 July 2023 (https://www.theguardian.com/world/2023/jul/11/uruguay-drought-water-google-data-center, accessed 12 July 2023).

116. Thornhill J. The sceptical case on generative AI. Financial Times, 17 August 2023. (https://www.ft.com/content/ed323f48-fe86-4d22-8151-eed15581c337, accessed 11 September 2023).

117. Marcus G. The imminent enshittification of the Internet. Substack, 16 August 2023 (https://garymarcus.substack.com/p/the-imminent-enshittification-of, accessed 11 September 2023).

118. Pause giant AI experiments: an open letter. Narberth (PA), Future of Life Institute, 2023 (https://futureoflife.org/open-letter/pause-giant-ai-experiments/, accessed 13 June 2023).

119. Perrigo B. DeepMind's CEO helped take AI mainstream. Now he's urging caution. Time, 12 January 2023 (https://time.com/6246119/demis-hassabis-deepmind-interview/, accessed 13 June 2023).

120. Lomas N. Unpacking the rules shaping generative AI. Tech Crunch, 13 April 2023 (https://techcrunch.com/2023/04/13/generative-ai-gdpr-enforcement/, accessed 13 June 2023).

121. Mökander J, Schuett J, Kirk HR, Floridi L. Auditing large language models: a three-layer approach. Soc Sci Res Netw 2023;DOI:10.2139/ssrn.4361607.

122. Lomas N. Report details how Big Tech is leaning on EU not to regulate general purpose Ais. Tech Crunch, 23 February 2023 (https://techcrunch.com/2023/02/23/eu-ai-act-lobbying-report/, accessed 20 June 2023).

123. Sambasivan N, Kapania S, Highfill H, Akrong D, Paritosh P, Aroyo LM. "Everyone wants to do model work, not the data work.": data cascades in high-stakes AI. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, May 2021;DOI:10.1145/3411764.3445518.

124. Browne G. AI is steeped in Big Tech's digital colonialism. Wired, 25 May 2023 (https://www.wired.co.uk/article/abeba-birhane-ai-datasets, accessed 17 June 2023).

125. Baxter K, Schelsinger, N. Managing the risks of generative AI. Harvard Business Review, 6 June 2023 (https://hbr.org/2023/06/managing-the-risks-of-generative-ai, accessed 17 June 2023).

126. Samuelson P. Generative AI meets copyright. Science 2023;381:158-61.

127. El-Mhamdi E, Farhadkhani S, Guerraoui R, Gupta N, Hoang L, Pinot R et al. On the impossible safety of large AI models. arXiv 2209.15259v2.

128 Open AI. GPT-4 technical report. ArXiv 2303.08774v3.

129. Murgia M. Open AI's red team: experts hired to "break" ChatGPT. Financial Times, 14 April 2023 (https://www.ft.com/content/0876687a-f8b7-4b39-b513-5fee942831e8, accessed 10 July 2023).

130. Clegg N. Openness on AI is the way forward for tech. Financial Times, 11 July 2023 (https://www.ft.com/content/ac3b585a-ce50-43d1-b71d-14dfe6dce999, accessed 11 July 2023).

131. Huang S, Toner H, Haluza Z, Creemers R, Webster G. Measures for the management of generative artificial intelligence services (draft for comment) (translation). DigiChina. Palo Alto (CA), Stanford University, Program on Geopolitics,

2023 (https://digichina.stanford.edu/work/translation-measures-for-the-management-of-generative-artificial-intelligence-services-draft-for-comment-april-2023/, accessed 17 June 2023).

132. Ye J. China says generative AI rules to apply only to products for the public. Reuters, 13 July 2023 (https://www.reuters.com/technology/china-issues-temporary-rules-generative-ai-services-2023-07-13/, accessed 13 July 2023).

133. Bommasani R, Klyman K, Zhang D, Liang P. Do foundation model providers comply with the draft EU AI Act? Palo Alto (CA), Stanford University, Human-centered Artifical Intelligence, 2021 (https://crfm.stanford.edu/2023/06/15/eu-ai-act.html, accessed 17 June 2023).

134. Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. Strasbourg, European Parliament, 2023 (https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html, accessed 10 July 2023).

135. Beyond ChatGPT: how can Europe become a leader in generative AI? Kaiserslautern, German Research Centre for Artificial Intelligence; 2023. (https://www.dfki.de/en/web/news/jenseits-von-chatgpt-wie-kann-europa-bei-der-generativen-ki-eine-fuehrungsposition-uebernehmen, accessed 17 June 2023).

136. Spirling A. Why open source generative AI models are an ethical way forward for science. Nature 2023;DOI:10.1038/d41586-023-01295-4.

137. Vincent J. Meta's powerful AI language model has leaked online - What happens now? The Verge, 8 March 2023 (https://www.theverge.com/2023/3/8/23629362/meta-ai-language-model-llama-leak-online-misuse, accessed 29 July 2023).

138. Maffuli S. Meta's Llama 2 license is not open source. Open Source Initiative, 20 July 2023 (https://opensource.org/blog/metas-llama-2-license-is-not-open-source, accessed 29 July 2023).

139. Marble A. Software licenses masquerading as open source. marble.onl, 1 June 2023 (https://www.marble.onl/posts/software-licenses-masquerading-as-open-source.html, accessed 29 July 2023).

140. Keary T. Report finds 82% of open source software components "inherently risky". Venture Beat, 17 April 2023 (https://venturebeat.com/security/report-finds-82-of-open-source-software-components-inherently-risky/, accessed 8 July 2023).

141. Generative AI raises competition concerns. Technology blog, 29 June 2023. Washington DC, Federal Trade Commission, 2023 (https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns, accessed 29 July 2023).

142. Wishart-Smith H. Generative AI: cybersecurity friend and foe. Forbes, 6 June 2023 (https://www.forbes.com/sites/heatherwishartsmith/2023/06/06/generative-ai-cybersecurity-friend-and-foe/?sh=4407e0884bd2, accessed 29 July 2023).

143. Metz C. Researchers poke holes in safety controls of ChatGPT and other chatbots. The New York Times, 27 July 2023 (https://www.nytimes.com/2023/07/27/business/ai-chatgpt-safety-research.html, accessed 11 September 2023).

144. Harris T, Freuh S. The complexity of technology's consequences is going up exponentially, but our wisdom and awareness are not. Issues in Science and Technology, 16 May 2023 (https://issues.org/tristan-harris-humane-technology-misinformation-ai-democracy/, accessed 19 June 2023).

145. Schyns C. The lobbying ghost in the machine: Big Tech's covert defanging of Europe's AI Act. Brussels, Corporate Europe Observatory; 2023 (https://corporateeurope.org/en/2023/02/lobbying-ghost-machine, accessed 17 June 2023).

146. Fact sheet: Biden-Harris Administration announces new actions to promote responsible AI innovation that protects Americans'rights and safety. Washington DC, White House, 4 May 2023 (https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/04/fact-sheet-biden-harris-administration-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety/, accessed 19 June 2023).

147. Sijbrandij Sid. AI weights are not "open source". Open Core Ventures, 27 June 2023. (https://opencoreventures.com/blog/2023-06-27-ai-weights-are-not-open-source, accessed 29 July 2023).

148. Meeker H. Towards an open weights definition. Copyleft Currents, 8 June 2023 (https://heathermeeker.com/2023/06/08/toward-an-open-weights-definition/, accessed 29 July 2023).

149. Dastin J, Tong A. Google, one of AI's biggest backers, warns its own staff about chatbots. Reuters, 15 June 2023 (https://www.reuters.com/technology/google-one-ais-biggest-backers-warns-own-staff-about-chatbots-2023-06-15/, accessed 9 July 2023).

150. Kanter GP, Packel EA. Health care privacy risks of AI chatbots. JAMA 2023;330:311-2. doi:10.1001/jama.2023.9618.

151. Interim measures for the management of generative artificial intelligence services. Beijing, Cyberspace Administration of China, 13 2023. (http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm, accessed 29 July 2023).

152. Satariano A. E.U. agrees on landmark artificial intelligence rules. The New York Times, 8 December 2023 (https://www.nytimes.com/2023/12/08/technology/eu-ai-act-regulation.html, accessed 15 December 2023).

153. The EU should regulate on the basis of rights, not risks. Access Now, 17 February 2021 (https://www.accessnow.org/eu-regulation-ai-risk-based-approach/, accessed 21 June 2023).

154. Marks M, Haupt CE. AI chatbots, health privacy, and challenges to HIPAA compliance JAMA. 2023;330:309-10.

155. Marcus G. Two models of AI oversight - and how things could go deeply wrong. Substack, 8 June 2023 (https://garymarcus.substack.com/p/two-models-of-ai-oversight-and-how, accessed 17 June 2023).

156. Kang C, Metz C. FTC opens investigation into ChatGPT maker over technology's potential harms. The New York Times, 13 July 2023 (https://www.nytimes.com/2023/07/13/technology/chatgpt-investigation-ftc-openai.html, accessed 29 July 2023).

157. Ordish J. Large language models and software as a medical device. MedRegs blogs, 3 March 2023 (https://medregs.blog.gov.uk/2023/03/03/large-language-models-and-software-as-a-medical-device/, accessed 19 June 2023).

158. Gilbert S, Harvey H, Melvin T, Vollebregt E, Wicks P. Large language model AI chatbots require approval as medical devices. Nat Med 2023;DOI:10.1038/s41591-023-02412-6.

159. Ghost in the machine: Addressing the harm of generative AI. Forbrukerradet. Oslo, Norwegian Concumer Council, 2023 (https://storage02.forbrukerradet.no/media/2023/06/generative-ai-rapport-2023.pdf, accessed 9 July 2023).

160. Mökander J, Floridi L. Ethics-base auditing to develop trustworthy AI. Minds & Machines, 2021;DOI:10.1007/s11023-021-09557-8.

161. Minssen T, Vayena E, Cohen IG. The challenges for regulating medical use of ChatGPT and other large language models. JAMA 2023;330:315-6.

162. Questions and answers: AI Liability Directive. Brussels, European Commission, 2022 (https://ec.europa.eu/commission/presscorner/detail/en/QANDA_22_5793, accessed 20 June 2023).

163. Duffourc MN, Gerke S. The proposed EU directives for AI liability leave worrying gaps likely to impact medical AI. NPJ Digit Med 2023;DOI:10.1038/s41746-023-00823-w.

164. Regulatory considerations on artificial intelligence for health. Geneva, World Health Organization, 2023 (https://iris.who.int/bitstream/handle/10665/373421/9789240078871-eng.pdf?sequence=1&isAllowed=y, accessed 16 November 2023).

165. Marcus G. Artificial Intelligence is stuck. Here's how to move it forward. The New York Times, 29 July 2017 (https://www.nytimes.com/2017/07/29/opinion/sunday/artificial-intelligence-is-stuck-heres-how-to-move-it-forward.html, accessed 20 June 2023).

166. Parker G. Rishi Sunak to lobby Joe Biden for UK "leadership" role in AI development. Financial Times, 5 June 2023 (https://www.ft.com/content/7c30ea28-2895-44c2-9a2d-c31ea7fa27e7, accessed 19 June 2023).

167. Hogarth I. We must slow down the race to god-like AI. Financial Times, 13 April 2023 (https://www.ft.com/content/03895dc4-a3b7-481e-95cc-336a524f2ac2, accessed 10 July 2023).

168. Blinken A, Raimondo G. To shape the future of AI, we must act quickly. Financial Times, 24 July 2023 (https://www.ft.com/content/eea999db-3441-45e1-a567-19dfa958dc8f, accessed 30 July 2023).

169. Networked, inclusive multilateralism can help overcome challenges of era, says Secretary General, opening general assembly session, United Nations, 17 September 2019. (https://press.un.org/en/2019/sgsm19746.doc.htm, accessed 18 September 2023).

Allegato

Metodi

Questa guida è stata sviluppata attraverso il consenso. L'OMS si è affidata a un gruppo di esperti sull'etica e la governance dell'IA per la salute. Si tratta di 20 esperti da tutte le aree geografiche dell'OMS che si sono incontrati ogni due settimane per 4 mesi. Il gruppo di esperti ha applicato all'uso emergente degli LMM nell'assistenza sanitarie e nella medicina il consenso di principi e raccomandazioni della guida precedentemente rilasciata dall'OMS sull'etica e la governance dell'IA per la salute.

Il gruppo di esperti ha per prima cosa mappato preliminarmente gli usi potenziali, i benefici e i rischi degli LMM per gli utenti finali. Il gruppo di esperti ha anche identificato i rischi per i sistemi sanitari con l'uso di tali sistemi di IA. Questa fase è stata integrata da una ricerca completa in letteratura riguardo a: usi esistenti e proposti degli LMM nell'assistenza sanitaria dopo anni della sua evoluzione e del suo utilizzo; gli usi previsti degli LMM; le critiche e le analisi degli LMM pubblicati prima del rilascio di questa guida.

Basandosi su una conoscenza condivisa dei benefici e dei rischi potenziali, il gruppo di esperti ha identificato un framework per affrontare varie sfide etiche e opportunità associate all'uso degli LMM. Il gruppo di esperti ha concordato che un approccio di "catena di valore" era adatto a rappresentare dove e come organizzare una governance appropriata, e quale attore o attori debbano essere considerati responsabili nel portare avanti misure rilevanti.

Sebbene la legislazione e le misure regolatorie esistenti o proposte in varie giurisdizioni non sono state riferite o utilizzate come strumenti per le raccomandazioni, il gruppo di esperti ha prodotto una raccomandazione che può essere applicabile in vari Paesi con diverse leggi. Il gruppo di esperti ha anche dibattuto le raccomandazioni per le aziende, per gli acquirenti di tali sistemi di IA e per gli utenti finali di LMM, specialmente gli erogatori di assistenza sanitaria e i pazienti.

L'OMS riconosce che questa guida avrà la necessità di essere rivista e aggiornata per assicurare che le conclusioni e le raccomandazioni del gruppo di esperti rimangano rilevanti e utili.