AEQUITAS

Regolamentazione dell'intelligenza artificiale nel settore sanitario pubblico per prevenire pregiudizi di genere e razziali

Un modello normativo per l'IA biomedica in linea con la Carta

Documento programmatico

Co-funded by
the European Union

Introduzione

L'intelligenza artificiale (IA) è sempre più integrata nei sistemi sanitari di tutta Europa, a supporto della diagnosi, del processo decisionale clinico, della pianificazione terapeutica e della previsione del rischio per i pazienti. Queste tecnologie offrono notevoli opportunità per migliorare l'efficienza e gli esiti dell'assistenza sanitaria. Allo stesso tempo, prove sempre più numerose dimostrano che **i sistemi di IA possono riprodurre o amplificare le disuguaglianze esistenti nell'assistenza sanitaria** .[1]

Diversi meccanismi possono portare a risultati distorti nel processo decisionale supportato dall'IA in ambito sanitario. Tra questi figurano squilibri nei dati di addestramento, errori di misurazione nelle variabili cliniche, indicatori proxy inappropriati e differenze tra gli ambienti di sviluppo e i contesti di implementazione nel mondo reale.[2] I pregiudizi possono insorgere anche quando i sistemi di IA vengono introdotti nei flussi di lavoro clinici senza un'adeguata formazione per gli operatori sanitari o senza sostanziali meccanismi di supervisione umana e di responsabilità per monitorarne le prestazioni su popolazioni di pazienti eterogenee.[3]

Inoltre, la ricerca empirica ha dimostrato che tali **distorsioni possono avere gravi conseguenze nei contesti clinici**. Alcuni studi hanno segnalato disparità persistenti nella diagnosi e nel trattamento delle principali patologie. Le donne e i pazienti appartenenti a minoranze razziali o etniche affetti da malattie cardiovascolari ricevono spesso diagnosi e gestione ritardate o meno intensive rispetto agli uomini e ai pazienti bianchi,[4] e le differenze razziali ed etniche influenzano la cura e gli esiti del diabete,[5] mentre i disturbi di salute mentale come la depressione sono sottodiagnosticati e sottotrattati

[1] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H. e Wang, Y., *"L'intelligenza artificiale in ambito sanitario: passato, presente e futuro"*, *Stroke and Vascular Neurology* 2(4): 230–243, 2017.

[2] Vokinger, K. N., Feuerriegel, S., & Kesselheim, A. S., *Ridurre i pregiudizi nell'apprendimento automatico in medicina*, *Communications Medicine* 1, 25, 2021.

[3] Panesar, S., Kliot, M., Parrish, R., Fernandez-Miranda, J., Cagle, Y., & Britz, G.W., *«Promesse e rischi dell'intelligenza artificiale in neurochirurgia», Neurosurgery* 87(1): 33–44, 2020.

[4] Balla, S., Gomez, S.E., & Rodriguez, F., *Disparità nell'assistenza cardiovascolare e negli esiti per le donne appartenenti a minoranze razziali/etniche*, *Current Treatment Options in Cardiovascular Medicine* 22(12):75, 2020.

[5] Young, C. e Myers, A. K., *Disparità razziali ed etniche nell'assistenza clinica e nella gestione del diabete: una revisione narrativa*, *Endocrine Practice* 29(4): 295–300, 2023.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

sia nelle donne che nelle popolazioni appartenenti a minoranze.[6] Un esempio ben documentato è uno studio esplorativo pubblicato sul **Journal of Medical Internet Research**, che ha rilevato disparità di prestazioni basate sul sesso negli algoritmi di apprendimento automatico utilizzati per la previsione delle malattie cardiache; in particolare, è emerso che le pazienti di sesso femminile erano costantemente sottorappresentate nei set di dati di addestramento e che i modelli mostravano tassi di falsi negativi più elevati per le donne, dimostrando come dati distorti possano contribuire a **prestazioni diagnostiche inique**.[7] Estendendo questa preoccupazione all'equità razziale, un altro studio ha rilevato che i modelli di IA addestrati a prevedere la gravità della depressione sulla base del linguaggio utilizzato sui social media hanno ottenuto buoni risultati per i partecipanti bianchi ma scarsi per quelli neri, rivelando disparità di prestazioni basate sulla razza che riflettono differenze nei modelli di espressione non colte dai modelli.[8]

Affrontare questi rischi di parzialità richiede **modelli normativi che traducano i principi etici e giuridici internazionali ed europei in meccanismi di governance concreti** che le istituzioni sanitarie possano attuare nella pratica. All'interno dell'Unione europea (UE), il panorama normativo si è evoluto rapidamente con l'adozione dell'Artificial Intelligence Act (AI Act), che stabilisce obblighi per i sistemi di IA ad alto rischio e per i soggetti che li implementano, come gli ospedali.[9] Tuttavia, l'applicazione di questi requisiti giuridici come procedure pratiche all'interno delle istituzioni sanitarie rimane una sfida significativa.

Il presente documento programmatico propone **un modello normativo per la governance dell'IA biomedica che sia in linea con l'acquis dell'UE**, compresa la Carta dei diritti fondamentali dell'Unione europea, e sia concepito per i sistemi sanitari pubblici e le organizzazioni della società civile (OSC). Il modello proposto si fonda su strumenti normativi europei, tra cui l'AI Act, il regolamento sui dispositivi medici (MDR), il regolamento sui dispositivi diagnostici in vitro (IVDR), il regolamento generale sulla protezione dei dati (GDPR) e l'emergente Spazio europeo dei dati sanitari (EHDS). Esso traduce i principi dei diritti fondamentali, in particolare la dignità umana, l'uguaglianza,

---

[6] Sorkin, D.H., Ngo-Metzger, Q., Billimek, J., August, K.J., Greenfield, S., & Kaplan, S.H., Depressione sottodiagnosticata e sottotrattata tra pazienti di diversa origine razziale/etnica con diabete di tipo 2, *Diabetes Care* 34(3): 598–600, 2011.

[7] Straw, I., Rees, G., & Nachev, P., *Disparità di prestazioni in base al sesso negli algoritmi di apprendimento automatico per la previsione delle malattie cardiache: studio esplorativo*, *Journal of Medical Internet Research* 26: e46936, 2024.

[8] Rai, S., Stade, E. C., Giorgi, S., Francisco, A., Ungar, L. H., Curtis, B., & Guntuku, S. C., *I principali indicatori linguistici della depressione sui social media dipendono dalla razza*, *Atti dell'Accademia Nazionale delle Scienze* 121(14): e2319837121, 2024.

[9] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, *che stabilisce norme armonizzate in materia di intelligenza artificiale (Legge sull'IA)*, PE/24/2024/REV/1, Gazzetta ufficiale dell'Unione europea, 13 giugno 2024.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

la non discriminazione e il diritto all'assistenza sanitaria, in garanzie pratiche applicabili all'intero ciclo di vita dei sistemi di IA. I riferimenti a esempi selezionati provenienti da giurisdizioni extra-UE sono inclusi solo laddove forniscano spunti trasferibili e siano intesi a integrare il quadro giuridico europeo basato sui diritti che sta alla base del modello.

Inoltre, il modello normativo attinge ad approcci normativi consolidati provenienti da altri settori ad alto rischio, tra cui la gestione della sicurezza aerea, la governance del rischio dei modelli finanziari, la farmacovigilanza nella regolamentazione dei medicinali e le valutazioni d'impatto algoritmico nella pubblica amministrazione. Questi approcci forniscono insegnamenti preziosi per la costruzione di quadri di governance in grado di affrontare i complessi rischi legati al ciclo di vita associati all'IA biomedica e sono stati qui adattati per integrare esplicitamente le considerazioni relative all'equità di genere e razziale in tutto il ciclo di vita dell'IA.

Il modello fornisce alle istituzioni sanitarie e alle organizzazioni della società civile (CSO) un **quadro di governance pratico che copre l'approvvigionamento, la convalida, l'implementazione, il monitoraggio e la responsabilità** per i sistemi di IA biomedica. Integrando la tutela dei diritti fondamentali nei processi operativi, questo quadro mira a ridurre i rischi di pregiudizi di genere e razziali nel processo decisionale sanitario supportato dall'IA e a contribuire a risultati sanitari più equi in tutta l'UE. Inoltre, il documento **offre una serie di politiche, strumenti e risorse pratiche che gli ospedali e le organizzazioni della società civile possono adattare ai diversi contesti nazionali** per attuare il modello in modo coerente in contesti reali.

Il quadro giuridico europeo per l'IA nel settore sanitario

L'UE ha istituito un contesto normativo a più livelli per l'IA biomedica, costituito da quattro livelli giuridici complementari: la legge sull'IA, la normativa sui dispositivi medici e le relative linee guida, la protezione dei dati e la governance dei dati sanitari.

L'AI Act istituisce un quadro normativo basato sul rischio per i sistemi di IA. Molte applicazioni sanitarie, compresi gli strumenti di supporto alle decisioni cliniche e gli algoritmi diagnostici, rientrano nella categoria dei sistemi di IA ad alto rischio, che devono soddisfare requisiti rigorosi in materia di gestione del rischio, governance dei dati, trasparenza, supervisione umana e accuratezza. I requisiti applicabili comprendono anche l'identificazione e la mitigazione dei rischi di esiti discriminatori attraverso un'analisi disaggregata delle popolazioni, considerando il sesso e il genere, l'origine etnica e altre categorie sociali come variabili trasversali. **Anche le istituzioni sanitarie che implementano sistemi di IA hanno obblighi specifici ai sensi della legge sull'IA**. Questi includono garantire che i sistemi siano utilizzati in conformità con le istruzioni del fornitore, monitorarne le prestazioni, mantenere registri operativi e segnalare incidenti

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

gravi alle autorità competenti, tra l'altro nei casi in cui tali incidenti possano influire in modo sproporzionato su diverse popolazioni di pazienti.

Per quanto riguarda l'applicazione della legge sull'IA, nel 2025 sono già entrate in vigore le pratiche vietate e sono state avviate alcune norme di supervisione e governance.[10] La prossima tappa fondamentale è l'agosto 2026, quando entrerà in vigore la maggior parte degli obblighi previsti dalla legge (compresi i sistemi ad alto rischio di cui all'Allegato III e gli obblighi di trasparenza di cui all'articolo 50). L'applicazione dell'IA ad alto rischio integrata in prodotti regolamentati (che comprende molti sistemi di IA per dispositivi medici) è prevista a partire dall'agosto 2027.

Per gli ospedali, le disposizioni della legge sull'IA più rilevanti dal punto di vista pratico sono:

- **Requisiti di sistema ad alto rischio**. I fornitori devono implementare un sistema di gestione dei rischi per l'intero ciclo di vita (art. 9) e pratiche di governance dei dati che prevedano: aspettative esplicite in materia di individuazione e mitigazione dei pregiudizi (art. 10), compresa la valutazione della rappresentatività dei dati e delle potenziali disparità di prestazioni tra le popolazioni mediante analisi disaggregate per sesso e genere e per origine razziale o etnica; garanzie di trasparenza e fornitura di informazioni agli utenti (art. 13); un controllo umano volto a prevenire i pregiudizi legati all'automazione e a consentire procedure di esclusione o interruzione (art. 14); e controlli di accuratezza, robustezza e sicurezza informatica con particolare attenzione ai circuiti di retroazione e agli attacchi specifici all'intelligenza artificiale (art. 15), che, se non affrontati, potrebbero altrimenti aggravare le disuguaglianze esistenti.

- **Obblighi dei soggetti che implementano i sistemi**. Gli ospedali e gli altri soggetti che implementano l'IA devono utilizzare i sistemi di IA ad alto rischio in linea con le istruzioni del fornitore (art. 26); assegnare una supervisione umana competente (art. 26); garantire che i dati di input siano rappresentativi e completi laddove il soggetto che implementa l'IA controlli l'immissione dei dati (art. 26), tenendo conto anche (ove fattibile) della diversità demografica nelle popolazioni cliniche; monitorare il funzionamento del sistema (art. 26); conservare i registri generati automaticamente per almeno sei mesi (art. 26); e segnalare tempestivamente gli incidenti gravi ai fornitori e alle autorità competenti (artt. 26, 73), in particolare laddove sussista il rischio di esiti clinici diseguali.

[10] Commissione europea, *Calendario per l'attuazione della legge UE sull'IA*, Bruxelles: Commissione europea, 2024.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

- **Valutazione d'impatto sui diritti fondamentali (FRIA)**. Prima di implementare determinati sistemi di IA ad alto rischio, gli enti di diritto pubblico (e i soggetti privati che forniscono servizi pubblici) devono effettuare una FRIA, descrivendo i processi, i gruppi interessati, i rischi, le misure di supervisione umana e i meccanismi di mitigazione o di reclamo, e successivamente notificarla all'autorità di vigilanza del mercato (art. 28). Questo requisito è direttamente rilevante per gli ospedali pubblici in molti Stati membri, poiché fornisce un meccanismo per identificare e affrontare, prima dell'implementazione, i rischi di discriminazione e gli impatti iniqui valutati attraverso un'analisi disaggregata per sesso e genere e per razza/etnia.

Oltre alla legge sull'IA, **i sistemi di IA utilizzati in contesti medici sono soggetti anche alla normativa sui dispositivi medici e alle relative linee guida**, tra cui il *regolamento sui dispositivi medici (MDR)*[11] e il *regolamento sui dispositivi diagnostici in vitro (IVDR),*[12] che disciplinano la sicurezza e le prestazioni delle tecnologie mediche. Tali regolamenti richiedono una valutazione clinica e una sorveglianza post-commercializzazione per il software dei dispositivi medici, compresi gli strumenti basati sull'IA, con particolare attenzione alle prestazioni eque tra le popolazioni di pazienti disaggregate, laddove i dati lo consentano. Inoltre, la linea guida del Gruppo di coordinamento sui dispositivi medici (MDCG) *sulla qualificazione e la classificazione del software ai sensi dell'MDR/IVDR* fornisce linee guida su come qualificare e classificare il software ai sensi dell'MDR e dell'IVDR.[13] Essa sottolinea che la classificazione dipende dalla destinazione d'uso piuttosto che dalla collocazione del software (ad esempio, basato su cloud o integrato) e include esempi tratti dai sistemi informativi ospedalieri e dal software di supporto alle decisioni cliniche, evidenziando la necessità di considerare gli impatti differenziali sulle diverse popolazioni di pazienti. *La Guida* dell'MDCG *sulla valutazione clinica e delle prestazioni del software per dispositivi medici* fornisce ulteriore supporto, rafforzando l'aspettativa che le prove siano proporzionate al ruolo clinico del software, comprese le prove che le prestazioni dell'IA siano coerenti tra le popolazioni differenziate per sesso e genere e per origine razziale o etnica.[14]

---

[11] Unione europea, *Regolamento (UE) 2017/745 sui dispositivi medici (MDR)*, Gazzetta ufficiale dell'Unione europea, L117, Bruxelles: UE, 2017.

[12] Unione europea, *Regolamento (UE) 2017/746 sui dispositivi medici diagnostici in vitro (IVDR)*, Gazzetta ufficiale dell'Unione europea, L117, Bruxelles: UE, 2017.

[13] Gruppo di coordinamento sui dispositivi medici, *Linee guida sulla qualificazione e la classificazione del software nel Regolamento (UE) 2017/745 (MDR) e nel Regolamento (UE) 2017/746 (IVDR)*, Bruxelles: MDCG, 2019.

[14] Gruppo di coordinamento sui dispositivi medici, *Linee guida sulla valutazione clinica (MDR) e sulla valutazione delle prestazioni (IVDR) del software nei dispositivi medici*, Bruxelles: MDCG, 2020.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Il rapporto tra l'AI Act e la normativa vigente sui dispositivi medici è chiarito nelle *Linee guida sull'interazione tra l'AI Act e l'MDR/IVDR*.[15] Il documento spiega che l'AI Act si applica parallelamente a questi due regolamenti. In pratica, ciò significa **che** molti **sistemi di IA medica devono conformarsi contemporaneamente a entrambi i quadri normativi**. La guida chiarisce inoltre quando l'IA medica è classificata come ad alto rischio ai sensi dell'AI Act, cosa che in genere si verifica quando il sistema di IA stesso è un dispositivo medico (o un componente di sicurezza di un dispositivo medico) e quindi soggetto a una valutazione di conformità da parte di un organismo notificato. Per gli appalti ospedalieri, ciò ha implicazioni pratiche. Gli ospedali dovrebbero richiedere ai fornitori di dimostrare lo stato normativo dei propri sistemi, compresa la classificazione MDR/IVDR, il coinvolgimento di un organismo notificato ove applicabile e l'allineamento ai requisiti relativi all'alto rischio previsti dalla legge sull'IA, nonché la documentazione relativa alle prestazioni disaggregate per gruppi di popolazione e alle misure di mitigazione dei pregiudizi, ove pertinente. I contratti di appalto dovrebbero inoltre richiedere ai fornitori di fornire la documentazione, le informazioni sulla trasparenza e le misure di supervisione necessarie agli enti che implementano i sistemi ai sensi della legge sull'IA, comprese disposizioni volte a garantire esiti equi per tutti i pazienti.

Anche diversi **altri strumenti dell'UE rivestono importanza per la governance dell'IA in ambito sanitario nella pratica**. Il *Regolamento generale sulla protezione dei dati (GDPR)* rimane un vincolo fondamentale per l'IA biomedica, poiché i dati sanitari sono dati personali di categoria speciale e le implementazioni ospedaliere richiedono spesso valutazioni d'impatto sulla protezione dei dati (DPIA). La legge sull'IA (artt. 26, paragrafo 9, e 27, paragrafo 4) collega esplicitamente la prassi degli operatori agli obblighi relativi alla DPIA (utilizzo delle informazioni fornite dal fornitore a supporto della DPIA) e posiziona la FRIA come complementare alla DPIA, la quale può includere la considerazione di impatti disparati tra popolazioni differenziate per sesso e genere e per origine razziale/etnica. Inoltre, lo *Spazio europeo dei dati sanitari (EHDS)* mira a facilitare l'accesso sicuro ai dati sanitari per l'erogazione dell'assistenza sanitaria e la ricerca, mantenendo al contempo solide garanzie di protezione dei dati.[16] Si prevede che l'EHDS svolga un ruolo chiave nel consentire la creazione di set di dati più rappresentativi per lo sviluppo dell'IA biomedica, sostenendo una valutazione incentrata sull'equità e riducendo il rischio di sottorappresentazione delle popolazioni di pazienti con determinate caratteristiche demografiche.[17]

---

[15] Gruppo di coordinamento sui dispositivi medici e Comitato europeo per l'intelligenza artificiale, *Linee guida sull'interazione tra MDR/IVDR e la legge sull'intelligenza artificiale*, Bruxelles: Commissione europea, 2025.

[16] Unione europea, *Regolamento sullo Spazio europeo dei dati sanitari (EHDS)*, Bruxelles: Unione europea, 2024.

[17] Commissione europea, *Spazio europeo dei dati sanitari (EHDS)*, Bruxelles: Commissione europea, 2025.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Infine, per quanto riguarda la sicurezza informatica, l'Agenzia dell'Unione europea per la sicurezza informatica ha elaborato linee guida sulle pratiche di sicurezza informatica per l'IA. In esse si sostiene che gli standard organizzativi esistenti (ad esempio, la gestione della sicurezza e della qualità secondo le norme ISO) possano essere adattati ai contesti dell'IA.[18] Le *Linee guida* del MDCG *sulla sicurezza informatica per i dispositivi medici* illustrano come i fabbricanti debbano soddisfare i requisiti relativi alla sicurezza informatica previsti dal MDR e dall'IVDR attraverso la gestione del rischio e il rispetto di requisiti minimi di sicurezza informatica.[19] È importante sottolineare che una solida governance della sicurezza informatica è alla base di un'implementazione equa dell'IA: proteggere l'integrità dei dati, l'accesso e l'affidabilità dei sistemi contribuisce a prevenire un'amplificazione involontaria delle disuguaglianze di genere e razziali/etniche.

Nel loro insieme, questi strumenti giuridici creano un contesto normativo complesso in cui devono operare i sistemi di IA nel settore sanitario. Il modello normativo sull'IA per gli ospedali pubblici europei qui proposto è progettato espressamente attorno a quei quattro «livelli» giuridici fondamentali che regolano la pratica dell'IA in ambito sanitario. Rispettivamente, il modello prevede meccanismi per monitorare e mitigare gli impatti iniqui sulle popolazioni differenziate per sesso e genere e per origine razziale o etnica durante l'intero ciclo di vita dell'IA.

Quadri di governance europei e internazionali applicabili all'IA responsabile nel settore sanitario

È importante notare che i quadri normativi e di standardizzazione europei sopra discussi costituiscono il fondamento giuridico primario per la governance dell'IA biomedica all'interno dell'Unione. Oltre al contesto normativo dell'UE per l'IA biomedica, diversi strumenti internazionali non vincolanti forniscono orientamenti per la gestione dei rischi legati all'IA. In particolare, i **quadri di governance aggiuntivi esaminati in questa sezione sono inclusi per illustrare tecniche di attuazione che possono essere trasferite nel contesto dell'UE, pur rimanendo pienamente subordinate ai requisiti giuridici europei e ai principi dei diritti fondamentali**. Essi vengono utilizzati in modo selettivo come analoghi provenienti da altre giurisdizioni in cui sono stati creati nella pratica meccanismi di governance del ciclo di vita comparabili, in particolare in settori quali la

[18] Agenzia dell'Unione europea per la sicurezza informatica (ENISA), *«Cybersecurity of AI and Standardisation»*, Atene: ENISA, 2023.

[19] Gruppo di coordinamento sui dispositivi medici, *MDCG 2019-16 Rev. 1 – Linee guida sulla sicurezza informatica dei dispositivi medici*, Bruxelles: Commissione europea, 2020.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

gestione del rischio dei modelli, l'audit algoritmico e la valutazione d'impatto istituzionale.

Ad esempio, il *quadro di riferimento per la gestione dei rischi dell'IA* (RMF) del National Institute of Standards and Technology (NIST) offre un approccio completo alla gestione dei rischi dell'IA attraverso un approccio basato sul ciclo di vita, organizzato attorno alla governance, alla mappatura dei rischi, alla misurazione dei rischi e alla gestione dei rischi.[20] Il quadro di riferimento riconosce esplicitamente l'equità e la lotta ai pregiudizi dannosi come dimensioni fondamentali di un'IA affidabile. Inoltre, nell'ambito della mappatura e della misurazione dei rischi, incoraggia le organizzazioni a identificare e valutare gli impatti disparati tra le diverse popolazioni di pazienti, tenendo conto anche delle differenze di sesso e genere e di quelle razziali/etniche. L'obiettivo è integrare l'affidabilità e i controlli dei rischi in tutte le fasi di progettazione, implementazione e monitoraggio, combinando la valutazione tecnica con la responsabilità organizzativa. Per gli ospedali, il punto di forza dell'AI RMF risiede nella sua capacità di tradursi in pratiche concrete, quali **ruoli definiti per la supervisione dell'equità, la valutazione metodica delle prestazioni disaggregate per popolazione e strategie di mitigazione documentate**, pur mantenendo una prospettiva più ampia basata sui diritti.

Per quanto riguarda i sistemi di gestione, la norma *ISO/IEC 42001* dell'Organizzazione internazionale per la normazione (ISO) definisce i requisiti per un sistema organizzativo di gestione dell'IA, concentrandosi su politiche, obiettivi, processi e miglioramento continuo per un'IA responsabile.[21] Pur non essendo prescrittiva in merito a danni specifici, la norma sostiene l'integrazione dei principi di equità e non discriminazione nelle politiche organizzative e nei controlli dei rischi, consentendo alle istituzioni di affrontare potenziali pregiudizi di genere e razziali in modo coordinato. A integrazione di ciò, *la norma ISO/IEC 23894* fornisce indicazioni sull'integrazione della gestione del rischio nello sviluppo e nell'implementazione dell'IA, compresa l'identificazione e la gestione dei rischi legati a impatti non intenzionali o sproporzionati su diversi gruppi di popolazione.[22] Per quanto riguarda specificamente l'IA clinica, *la norma ISO 14971* definisce i principi di gestione del rischio ampiamente utilizzati nella regolamentazione dei dispositivi medici, che descrivono l'identificazione sistematica dei pericoli, la valutazione del rischio, il controllo del rischio e il monitoraggio del ciclo di vita.[23] Questi processi possono essere estesi per tenere conto di prestazioni o esiti iniqui tra

[20] Istituto nazionale di standard e tecnologia, *Quadro di riferimento per la gestione dei rischi dell'intelligenza artificiale (AI RMF 1.0)*, Gaithersburg, MD: NIST, 2023.

[21] Organizzazione internazionale per la normazione, *ISO/IEC 42001:2023 – Sistema di gestione dell'intelligenza artificiale*, Ginevra: ISO, 2023.

[22] Organizzazione internazionale per la normazione, *ISO/IEC 23894:2023 – Gestione dei rischi legati all'intelligenza artificiale*, Ginevra: ISO, 2023.

[23] Organizzazione internazionale per la normazione, *ISO 14971:2019 – Dispositivi medici: applicazione della gestione dei rischi ai dispositivi medici*, Ginevra: ISO, 2019.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

popolazioni disaggregate nell'ambito dell'identificazione dei danni. Le norme descritte sono preziose in quanto suggeriscono **una struttura di sistema di gestione che gli ospedali possono adottare**. Forniscono inoltre un linguaggio comune attraverso il quale i requisiti relativi all'equità, quali la valutazione delle popolazioni disaggregate e la mitigazione dei pregiudizi, possono essere specificati negli appalti e nei contratti con i fornitori.

Anche i quadri di governance etica forniscono importanti orientamenti normativi. Le *Linee guida etiche* della Commissione europea *per un'IA affidabile* sono uno strumento politico utile, sebbene non vincolante.[24] Esse definiscono sette requisiti fondamentali per sistemi di IA responsabili, tra cui diversità, non discriminazione ed equità, che richiedono esplicitamente di evitare i pregiudizi e di considerare gli impatti su popolazioni con caratteristiche demografiche diverse, quali sesso e genere e origine razziale/etnica. *L'Elenco di valutazione per un'intelligenza artificiale affidabile (ALTAI)*, uno strumento volontario di autovalutazione sviluppato dal Gruppo di esperti di alto livello sulla intelligenza artificiale della Commissione europea, attua tali linee guida sotto forma di elenco di autovalutazione.[25] In un modello normativo per l'assistenza sanitaria, il valore di ALTAI risiede nel suo **formato basato su domande, che può essere adattato in pratiche liste di controllo per l'approvvigionamento e l'implementazione**, spingendo le parti interessate a considerare questioni quali la rappresentatività dei dati di addestramento, il rischio di esiti clinici disparati e l'adeguatezza delle misure di salvaguardia contro gli effetti discriminatori, integrando così le considerazioni di equità nel processo decisionale di routine da parte di medici, specialisti IT e dirigenti ospedalieri.

Da una prospettiva dei diritti umani, la *Convenzione quadro* del Consiglio d'Europa *sull'IA* è uno strumento giuridico rilevante, esplicitamente ancorato ai principi di dignità umana, trasparenza, responsabilità, uguaglianza, non discriminazione e privacy, fornendo così una base giuridica per affrontare gli effetti discriminatori, compresi quelli legati ai pregiudizi di genere e razziali nei sistemi di IA.[26] Analogamente, i *Principi sull'intelligenza artificiale* dell'Organizzazione per la cooperazione e lo sviluppo economico[27] e *la Raccomandazione* dell'UNESCO *sull'etica dell'IA*[28] offrono orientamenti normativi internazionali che incoraggiano l'identificazione e la mitigazione

[24] Commissione europea, *Linee guida etiche per un'IA affidabile*, Bruxelles: Commissione europea, 2019.

[25] Commissione europea, *Elenco di valutazione per un'intelligenza artificiale affidabile (ALTAI) – Autovalutazione*, Bruxelles: Commissione europea, 2020.

[26] Consiglio d'Europa, *Convenzione quadro sull'intelligenza artificiale (CAI)*, Strasburgo: Consiglio d'Europa, 2023.

[27] Organizzazione per la cooperazione e lo sviluppo economico, *Principi dell'OCSE sull'intelligenza artificiale*, Parigi: OCSE, 2019.

[28] UNESCO, *Raccomandazione sull'etica dell'intelligenza artificiale*, Parigi: UNESCO, 2021.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

di impatti ingiusti o disparati su popolazioni disaggregate, rafforzando l'aspettativa che i sistemi di IA non debbano riprodurre o amplificare le disuguaglianze strutturali. Per quanto riguarda la governance etica specifica nel settore sanitario, l'Organizzazione Mondiale della Sanità ha articolato sei principi fondamentali per l'IA in ambito sanitario: autonomia, benessere, spiegabilità, responsabilità, equità e reattività.[29] Il principio di **equità richiede esplicitamente di prestare attenzione alle differenze in termini di accesso, prestazioni e risultati tra le diverse popolazioni di pazienti, tenendo conto anche delle differenze di sesso e genere, nonché di quelle razziali o etniche**.

Questi quadri di governance internazionali si allineano direttamente ai requisiti della Carta in materia di dignità, uguaglianza, non discriminazione e diritto alla salute, e condividono un'enfasi comune sulla governance del ciclo di vita – principi che possono essere tradotti direttamente in controlli pratici per le istituzioni sanitarie.

Insegnamenti tratti dai modelli normativi intersettoriali

Settori ad alto rischio come quello aeronautico, bancario e farmaceutico hanno da tempo sviluppato sistemi di governance sofisticati per la gestione dei rischi tecnologici. Questi settori offrono preziosi approcci normativi che possono essere adattati alla governance dell'IA in ambito sanitario e alla gestione delle relative sfide legate ai pregiudizi di genere e razziali.

Sebbene i pregiudizi di genere e razziali non siano rischi tecnologici in senso stretto, essi emergono come rischi socio-tecnici quando dati, algoritmi e pratiche organizzative interagiscono con le disuguaglianze strutturali nel settore sanitario. In questo contesto, il termine «rischio» si riferisce alla manifestazione a livello di sistema dei pregiudizi nei risultati, nelle decisioni e negli esiti clinici, piuttosto che alla loro origine. Il settore sanitario stesso fornisce già ampie prove di tali effetti, comprese le disparità documentate nella diagnosi, nel trattamento e nella previsione del rischio tra popolazioni differenziate per sesso, genere e razza/etnia. Gli esempi intersettoriali integrano quindi le prove provenienti dal settore sanitario, illustrando meccanismi di governance consolidati per la gestione di sistemi complessi e ad alto impatto, insieme a risultati empirici provenienti dal settore sanitario che dimostrano la necessità di tali controlli normativi.

Un esempio particolarmente rilevante è il *Sistema di Gestione della Sicurezza (SMS)* utilizzato nel settore dell'aviazione. Le normative internazionali sull'aviazione civile trattano la sicurezza non come un esercizio di certificazione una tantum, ma come un processo continuo a livello organizzativo che consente di prendere decisioni basate

[29] Organizzazione Mondiale della Sanità, *Etica e governance dell'intelligenza artificiale per la salute*, Ginevra: OMS, 2021.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

sull'identificazione dei pericoli, la gestione dei rischi, la garanzia della sicurezza (monitoraggio e audit), la gestione del cambiamento e il miglioramento continuo.[30] Applicare questa logica all'IA nel settore sanitario significherebbe **istituire un «Sistema di gestione della sicurezza dell'IA» che incorpori esplicitamente considerazioni di equità**, tra cui: (1) un canale di segnalazione di pericoli e quasi-incidenti per i medici che rilevi potenziali episodi di pregiudizio di genere o razziale; (2) un registro formale dei rischi per ciascun sistema di IA che documenti le differenze di prestazioni tra popolazioni di pazienti disaggregate; (3) meccanismi di controllo delle modifiche che valutino l'impatto degli aggiornamenti dei modelli, della deriva dei dati o delle modifiche al flusso di lavoro sui risultati in termini di equità; e (4) revisioni periodiche di garanzia della sicurezza con azioni correttive che affrontino sia i rischi tecnici sia le disuguaglianze di genere o razziali identificate. L'enfasi esplicita posta dall'AI Act (artt. 9, 72) sulla gestione dei rischi durante l'intero ciclo di vita e sul monitoraggio post-commercializzazione è strettamente in linea con questa logica del SMS.

Un secondo approccio utile proviene dalla regolamentazione finanziaria, dove gli strumenti quantitativi complessi sono considerati "modelli" che richiedono uno sviluppo solido, una validazione indipendente e una supervisione da parte degli organi di governance. Nella vigilanza bancaria, quadri di governance del rischio analoghi sono stati implementati in giurisdizioni quali gli Stati Uniti (ad esempio, *le Linee guida di vigilanza* della Federal Reserve statunitense *sulla gestione del rischio di modello)* e forniscono un esempio di come possano essere attuati la validazione, la supervisione e la revisione indipendente dei modelli.[31] Approcci simili potrebbero essere applicati ai sistemi di IA biomedica per garantire che **gli strumenti di supporto alle decisioni cliniche siano convalidati in modo indipendente (internamente o esternamente) prima di essere introdotti** in contesti sanitari, con **una valutazione esplicita delle prestazioni nelle diverse popolazioni differenziate per sesso e genere, nonché per caratteristiche razziali o etniche**. I documenti minimi di convalida dovrebbero documentare lo scopo previsto del sistema, i limiti operativi, le modalità di guasto note e le prestazioni disaggregate per popolazione, comprese eventuali disparità o misure di mitigazione applicate. La riconvalida dovrebbe essere avviata ogni volta che il sistema viene implementato in una nuova sede, applicato a una popolazione di pazienti diversa, collegato a nuove apparecchiature o protocolli, aggiornato dal fornitore o mostra segni di deriva delle prestazioni. Queste pratiche sono coerenti con l'AI Act (artt. 13, 14, 60), che impone test, trasparenza nei confronti di chi implementa il sistema e supervisione

[30] Organizzazione internazionale dell'aviazione civile, *Manuale di gestione della sicurezza (Doc 9859)*, 4ª ed., Montréal: ICAO, 2018.

[31] Consiglio dei governatori del Sistema della Riserva Federale e Ufficio del Controllore della Valuta, *Linee guida di vigilanza sulla gestione del rischio di modello (SR 11-7)*, Washington, D.C., 4 aprile 2011.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

umana volta a individuare e prevenire i pregiudizi dell'automazione, garantendo che l'assistenza supportata dall'IA sia sicura ed equa per tutte le diverse popolazioni.

Per quanto riguarda la governance post-implementazione, la regolamentazione farmaceutica offre un modello solido: nella farmacovigilanza, le reazioni avverse e i segnali di sicurezza vengono sistematicamente raccolti e analizzati per individuare i rischi emergenti. In particolare, *il Modulo VI delle Buone Pratiche di Farmacovigilanza (GVP)* dell'Agenzia europea per i medicinali definisce aspettative dettagliate per la raccolta, la gestione e la presentazione delle segnalazioni di sospette reazioni avverse, riflettendo un approccio ecosistemico alla segnalazione degli incidenti e al loro follow-up.[32] Un **sistema** analogo **di «vigilanza sull'IA» potrebbe consentire agli ospedali di segnalare e analizzare gli incidenti** relativi a decisioni cliniche supportate dall'IA, con particolare attenzione agli esiti in termini di equità. Gli elementi chiave includerebbero: (1) una segnalazione sistematica che rilevi eventuali disuguaglianze osservate nell'assistenza o negli esiti in base al sesso e al genere o alle differenze razziali/etniche; (2) la documentazione obbligatoria degli incidenti gravi in cui un pregiudizio possa aver contribuito a un trattamento differenziale; (3) l'obbligo di segnalare al fornitore e alle autorità competenti le preoccupazioni relative all'equità (informando le autorità qualora i pregiudizi causino danni gravi, quali diagnosi errate o trattamenti iniqui, mentre le disuguaglianze di routine rimangano nell'ambito del monitoraggio interno della qualità); e (4) funzioni di rilevamento di tendenze e segnali che identifichino modelli emergenti di pregiudizio tra reparti, strutture o nella rete sanitaria più ampia. Tale monitoraggio consentirebbe di adottare misure di mitigazione proattive e di perseguire un miglioramento continuo, garantendo che l'assistenza supportata dall'IA sia non solo sicura ma anche equa.

Infine, gli strumenti di governance del settore pubblico, quali le valutazioni d'impatto algoritmico e gli audit sui pregiudizi, offrono metodi pratici per valutare l'impatto sociale dei sistemi decisionali automatizzati, comprese le potenziali disparità legate al sesso e al genere o alle differenze razziali ed etniche. Questi strumenti traducono i principi etici e gli standard dei diritti umani in formati attuabili che le organizzazioni possono utilizzare per anticipare e mitigare le disuguaglianze prima dell'implementazione. Ad esempio, *la Valutazione dell'impatto algoritmico* federale del Canada assegna un punteggio ai rischi dei sistemi di IA e li collega a misure di mitigazione.[33] Questo strumento inquadra esplicitamente l'implementazione come un problema di governance orientato all'impatto. Inoltre, l'approccio della città di New York agli strumenti automatizzati per le decisioni in materia di occupazione richiede audit sui pregiudizi e obblighi di

[32] Agenzia europea per i medicinali, *GVP Modulo VI: Raccolta e segnalazione di sospette reazioni avverse*, Londra: EMA, 2017

[33] Segretariato del Consiglio del Tesoro del Canada, *Valutazione d'impatto algoritmica*, Ottawa: Governo del Canada, 2026.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

divulgazione nei confronti del pubblico.[34] Di conseguenza, gli ospedali e le organizzazioni della società civile (CSO) operanti nel settore sanitario potrebbero implementare un **pacchetto di "Valutazione dell'impatto dell'IA e audit" che combini un questionario strutturato incentrato sull'equità con audit periodici indipendenti sull'equità**, producendo una sintesi pubblicabile delle prestazioni tra popolazioni disaggregate e degli sforzi di mitigazione. L'implementazione di tale pacchetto contribuirebbe a promuovere la trasparenza, la responsabilità pubblica e la fiducia dei pazienti nell'assistenza supportata dall'IA.

L'inclusione di modelli normativi provenienti da settori quali la sicurezza aerea, la gestione del rischio dei modelli finanziari, la farmacovigilanza e la governance algoritmica è intesa come un trasferimento deliberato di principi di governance consolidati. Questi ambiti sono stati selezionati perché attuano concretamente approcci formalizzati all'individuazione, al monitoraggio e alla mitigazione dei rischi in sistemi complessi in cui possono verificarsi impatti differenziali sugli individui. Molti di essi incorporano già pratiche di equità a livello operativo, tra cui la stratificazione demografica del rischio nella farmacovigilanza, la convalida dei modelli e i controlli sui pregiudizi nella vigilanza finanziaria, nonché la segnalazione sistematica di eventi avversi che interessano popolazioni specifiche. La loro rilevanza risiede sia nella somiglianza concettuale, sia nel loro comprovato ricorso alla governance del ciclo di vita, alla segnalazione obbligatoria e a meccanismi di supervisione indipendenti che possono essere direttamente adattati per affrontare i rischi legati all'equità di genere e razziale nei sistemi di IA biomedica.

Progetto di modello normativo sull'IA, in linea con la Carta, per ospedali pubblici e organizzazioni della società civile

Questa sezione delinea un'architettura di modello normativo che si **fonda sui principi della Carta dell'UE, ma si concretizza in procedure pratiche che gli ospedali e le organizzazioni della società civile possono attuare** nella pratica. La Carta fornisce il fondamento normativo per la governance delle tecnologie sanitarie in Europa, in particolare attraverso i principi di dignità umana, uguaglianza, non discriminazione e diritto all'assistenza sanitaria.[35] Questi diritti sono particolarmente rilevanti quando i sistemi automatizzati influenzano le decisioni cliniche che incidono sulla cura dei pazienti, comprese le potenziali disparità tra popolazioni differenziate per sesso e genere e per origine razziale/etnica.

---

[34] Dipartimento per la tutela dei consumatori e dei lavoratori della città di New York, *Strumenti automatizzati per le decisioni in materia di occupazione*, New York: Città di New York, 2023.

[35] Unione Europea, *Carta dei diritti fondamentali dell'Unione Europea*, 2012/C 326/02, Strasburgo: UE.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

A livello più fondamentale, l'articolo 1 della Carta, che sancisce l'inviolabilità della dignità umana, sostiene i requisiti di governance volti a garantire che gli strumenti di IA siano utilizzati in modo da preservare il giudizio umano e rispettare l'autonomia del paziente. In termini pratici, questo principio richiede **misure di salvaguardia contro l'eccessiva dipendenza dalle raccomandazioni automatizzate, meccanismi per contestare le decisioni assistite dall'IA e strutture di supervisione clinica che prevengano il «bias di automazione»** – la tendenza dei decisori umani a fidarsi eccessivamente dei risultati algoritmici, che può amplificare le disuguaglianze se le raccomandazioni dell'IA svantaggiano sistematicamente determinate popolazioni disaggregate.

Analogamente, gli articoli 20 e 21, che sanciscono l'uguaglianza davanti alla legge e vietano la discriminazione, forniscono il fondamento giuridico per obblighi espliciti di equità e di mitigazione dei pregiudizi nell'IA biomedica. Gli ospedali che implementano sistemi di IA dovrebbero quindi attuare **processi per monitorare le prestazioni del sistema su diverse popolazioni di pazienti**, tenendo conto di sesso e genere, razza o etnia e combinazioni intersezionali di queste caratteristiche. Laddove vengano rilevate disuguaglianze nelle prestazioni o negli esiti, dovrebbero essere attuate azioni correttive attraverso l'adeguamento dei modelli, modifiche al flusso di lavoro clinico o percorsi decisionali alternativi, garantendo che l'implementazione dell'IA non consolidi le disuguaglianze strutturali.

Inoltre, l'articolo 35 della Carta garantisce il diritto a un elevato livello di protezione della salute umana e sostiene quindi ulteriormente il principio secondo cui i sistemi di IA possono essere implementati solo laddove contribuiscano in modo dimostrabile a un'assistenza sanitaria sicura, efficace ed equa. Le tecnologie di IA non dovrebbero introdurre nuove barriere all'assistenza né esacerbare le disparità esistenti che colpiscono popolazioni di pazienti storicamente poco servite o comunque svantaggiate, e il loro impatto su tali popolazioni di pazienti dovrebbe essere valutato continuamente.

Per garantire tali diritti nella pratica, il modello normativo proposto nel presente documento programmatico adotta un **approccio di governance basato sul ciclo di vita che struttura la supervisione dell'IA attorno a una serie di punti decisionali obbligatori o "gate**". Questi gate riflettono la logica normativa insita nell'AI Act dell'UE e nella relativa legislazione sanitaria, fornendo al contempo agli ospedali un percorso di governance chiaro e verificabile. Il modello presuppone che l'ospedale sia solitamente un «utilizzatore» (ai sensi dell'AI Act) e un «fornitore di assistenza sanitaria» ai sensi della legislazione sanitaria nazionale, ma contempla anche il caso in cui l'ospedale diventi un «fornitore/produttore» (come nell'esempio dell'IA sviluppata internamente).

Gli ospedali sono i principali utilizzatori e responsabili dell'implementazione dei sistemi di IA, mentre gli organismi di supervisione (CSO) agiscono come organismi indipendenti

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

di supervisione, audit e monitoraggio basati sui diritti, senza esercitare un controllo quotidiano sulle decisioni relative all'utilizzo clinico.

L'attuazione di questo modello normativo segue un principio di proporzionalità che riconosce che **gli ospedali e le organizzazioni della società civile (CSO) possono avere livelli diversi di capacità tecnica, organizzativa e analitica nei vari Stati membri e sistemi sanitari**. I requisiti minimi obbligatori si applicano in modo uniforme per garantire garanzie di base in materia di sicurezza, trasparenza ed equità. Inoltre, pratiche avanzate, quali l'analisi estesa e disaggregata della popolazione, l'audit esterno e sofisticati strumenti di monitoraggio dei pregiudizi, possono essere implementate progressivamente a seconda della capacità istituzionale. Anche l'allegato tecnico segue questo approccio, distinguendo tra requisiti minimi e requisiti raccomandati.

1. **Ambito di applicazione e fase di classificazione.** Prima dell'acquisto o dello sviluppo, le istituzioni devono determinare lo status normativo di un determinato strumento di IA. Ciò include la valutazione se il sistema sia qualificabile come software per dispositivi medici ai sensi dell'MDR o dell'IVDR, se costituisca un sistema di IA ad alto rischio ai sensi dell'AI Act, o se rientri in una categoria a rischio inferiore che richieda comunque una supervisione di governance. In questa fase iniziale, gli ospedali possono segnalare gli strumenti il cui uso previsto potrebbe avere implicazioni più rilevanti in termini di equità o influire su popolazioni di pazienti eterogenee, tenendo presente che tali strumenti richiederanno una valutazione più dettagliata dei pregiudizi di genere e razziali nelle fasi successive. Le linee guida elaborate dall'MDCG sulla qualificazione e la classificazione del software forniscono esempi pratici che possono aiutare a evitare errori di classificazione normativa e a preparare il terreno per le valutazioni relative ai pregiudizi e all'equità.[36] Gli ospedali guidano la classificazione e la determinazione normativa. I CSO forniscono consulenza indipendente sui rischi legati all'equità, ma in questa fase non hanno potere decisionale.
2. **Fase preliminare all'appalto e alla stipula del contratto.** Gli ospedali esercitano spesso la loro maggiore influenza sulla governance dell'IA durante la fase di appalto. In questa fase, le istituzioni dovrebbero garantire che i fornitori si impegnino contrattualmente ad affrontare i rischi di equità, inclusi i pregiudizi di genere e razziali, durante l'intero ciclo di vita dell'IA. Ai fornitori dovrebbe essere richiesto di fornire documentazione che dimostri la conformità normativa,

[36] Gruppo di coordinamento sui dispositivi medici, *Linee guida sulla qualificazione e la classificazione del software nel Regolamento (UE) 2017/745 (MDR) e nel Regolamento (UE) 2017/746 (IVDR)*, Bruxelles: MDCG, 2019.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

comprese le prove delle valutazioni di conformità, ove pertinente, i risultati della validazione delle prestazioni e le informazioni necessarie affinché gli enti che implementano il sistema possano adempiere ai propri obblighi ai sensi dell'AI Act. I contratti di appalto dovrebbero ripartire esplicitamente le responsabilità tra fornitore e utilizzatore, inclusi i requisiti relativi a trasparenza, documentazione, registrazione, supervisione umana e supporto post-implementazione. L'obiettivo in questa fase è integrare la consapevolezza dei pregiudizi e la responsabilità negli obblighi contrattuali, creando le basi per un'implementazione sicura ed equa. Gli ospedali sono responsabili delle decisioni relative agli appalti e degli accordi contrattuali. Le organizzazioni della società civile (CSO) possono esaminare la documentazione relativa agli appalti per valutare aspetti di equità e diritti ed emettere raccomandazioni non vincolanti.

3. **Fase di validazione.** La terza fase è quella di validazione, che garantisce che gli strumenti di IA siano valutati in modo indipendente prima di entrare nei flussi di lavoro clinici. La validazione dovrebbe valutare sia le prestazioni complessive sia i risultati su popolazioni di pazienti disaggregate, con particolare attenzione al sesso e al genere e all'origine razziale/etnica. Questo passaggio è fondamentale perché i sistemi di IA biomedica possono mostrare prestazioni disparate tra le diverse popolazioni quando i dati di addestramento non sono rappresentativi o quando vengono trascurate le differenze contestuali tra gli ambienti di sviluppo e di implementazione.[37] La valutazione dovrebbe inoltre documentare eventuali misure di mitigazione applicate ed evidenziare i rischi residui in materia di equità, fornendo agli ospedali le prove necessarie per prendere decisioni informate sull'implementazione. Gli ospedali sono responsabili di condurre o commissionare la validazione e di interpretarne i risultati ai fini delle decisioni di implementazione. I CSO forniscono una revisione indipendente delle prestazioni disaggregate e dei rischi di distorsione, senza esercitare alcun controllo sui risultati della validazione.
4. **Fase di implementazione.** Una volta che un sistema ha superato la validazione, la quarta fase è quella dell'implementazione. Gli ospedali devono garantire che le condizioni operative supportino un uso sicuro, equo e responsabile. Ciò include l'assegnazione di personale di supervisione qualificato, l'implementazione di meccanismi che consentano ai medici di ignorare o disabilitare le raccomandazioni algoritmiche e la garanzia che gli utenti comprendano lo scopo previsto del sistema, i suoi limiti e qualsiasi considerazione relativa alle prestazioni disaggregate. I medici dovrebbero essere informati di potenziali variazioni ingiuste nei risultati quando si tengono in

[37] Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H., *Garantire l'equità nell'apprendimento automatico per promuovere l'equità sanitaria*, *Annals of Internal Medicine*, 169(12), 866-872, 2018.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

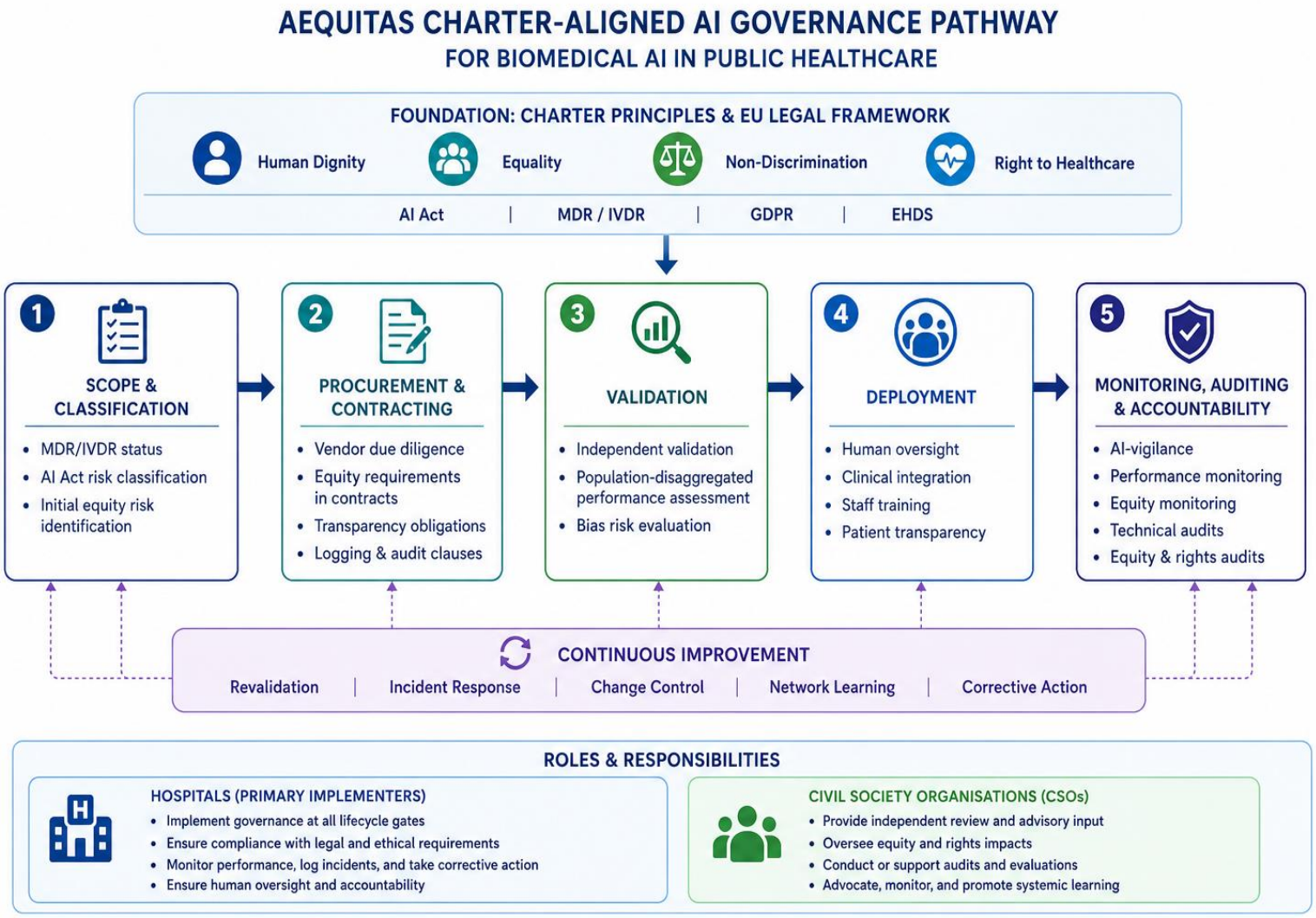
considerazione le differenze di sesso e genere o razziali/etniche, e i pazienti dovrebbero essere informati quando vengono utilizzati strumenti decisionali supportati dall'IA nella loro cura. Gli ospedali mantengono la piena responsabilità per le decisioni relative all'implementazione, all'integrazione clinica e alla supervisione dell'attuazione. I CSO non hanno alcun ruolo nell'implementazione, ma possono esaminare la documentazione relativa alla prontezza all'implementazione dal punto di vista dell'equità e dei diritti.

5. **Fase di monitoraggio, verifica e rendicontazione.** Questa fase finale regola la fase post-implementazione del ciclo di vita dell'IA. I sistemi di IA dovrebbero essere soggetti a un monitoraggio continuo delle prestazioni complessive, delle disparità tra le popolazioni disaggregate e degli incidenti relativi alla sicurezza o all'equità. Gli ospedali dovrebbero mantenere registri del funzionamento del sistema, implementare procedure standardizzate di segnalazione degli incidenti che rilevino potenziali pregiudizi di genere o razziali e rivalutare periodicamente le prestazioni del sistema attraverso esercizi di rivalidazione. Laddove emergano disuguaglianze o rischi di pregiudizio che non possano essere mitigati efficacemente, il sistema dovrebbe essere modificato, sospeso o dismesso. I risultati dovrebbero informare sia la governance interna sia, ove opportuno, un apprendimento più ampio a livello di rete per prevenire il ripetersi di pregiudizi. Gli ospedali sono responsabili del monitoraggio continuo, della registrazione, della segnalazione degli incidenti e dell'attuazione delle azioni di mitigazione. Le organizzazioni della società civile (CSO) conducono audit indipendenti, valutano i rischi sistemici di equità tra le diverse sedi e sostengono l'escalation delle disparità persistenti, senza esercitare alcun controllo operativo sulle modifiche al sistema.

Insieme, queste fasi del ciclo di vita trasformano i principi fondamentali dei diritti in processi di governance concreti in grado di affrontare i rischi di parzialità di genere e razziale nei sistemi di IA clinica.

Figura 1: Percorso di governance dell'IA allineato alla Carta AEQUITAS

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI



Politiche pratiche, strumenti operativi e strumenti di governance per i professionisti ospedalieri e delle organizzazioni della società civile

Per funzionare efficacemente in contesti sanitari reali, il modello normativo proposto in questo documento programmatico prevede la creazione di strumenti pratici che possano essere riutilizzati tra le istituzioni e adattati ai diversi contesti nazionali. Esso sottolinea la necessità di procedure, documentazione e modelli standardizzati che consentano sia agli ospedali che alle organizzazioni della società civile di gestire i sistemi di IA in modo coerente. La seguente serie di strumenti è in linea con i requisiti dell'AI

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS

Act, con le linee guida relative all'MDR e all'IVDR, e con le pratiche di governance intersettoriali.

Fase di appalto e contrattazione

- **Pacchetto per gli appalti e la stipula dei contratti.** La comunità dell'UE per gli appalti nel settore dell'IA ha pubblicato clausole contrattuali modello (versioni ad alto rischio e non ad alto rischio) e le ha aggiornate per riflettere l'AI Act, incoraggiando esplicitamente le autorità pubbliche a utilizzarle e fornendo commenti per la loro attuazione.[38] Per l'IA nel settore sanitario, al fine di valutare e mitigare i rischi di pregiudizi di genere e razziali, tali clausole possono essere adattate in un allegato specifico per gli ospedali, come documento autonomo, che richieda (come minimo):
    - (1) la rendicontazione standardizzata della composizione dei dati di addestramento, compresa la rappresentanza in termini di sesso, genere e razza/etnia, a supporto della valutazione della rappresentatività e del rischio di pregiudizio;
    - (2) la divulgazione della provenienza dei dati, specificando se i set di dati sono proprietari o non proprietari, le relative condizioni di accessibilità e le implicazioni in termini di trasparenza, validazione indipendente e valutazione dei pregiudizi di genere e razziali;
    - (3) la divulgazione delle prestazioni nelle diverse popolazioni disaggregate e delle prove di validazione, compresi i risultati stratificati per sesso e genere e per origine razziale o etnica, la documentazione di eventuali disparità note, le misure di mitigazione dei pregiudizi applicate durante lo sviluppo e i test di validazione a supporto (comprese le considerazioni specifiche del sito relative all'equità e alla correttezza);
    - (4) requisiti di registrazione e verificabilità, che garantiscano che il sistema sia in grado di generare registri di utilizzo in grado di supportare verifiche indipendenti relative a pregiudizi o impatti disparati;
    - (5) obblighi di aggiornamento e controllo delle modifiche, che impongono al fornitore di informare l'ospedale in merito ad aggiornamenti del modello, corsi di aggiornamento o modifiche al flusso di lavoro, e di documentarne i potenziali effetti sulle prestazioni in relazione alle popolazioni differenziate;
    - (6) impegni in materia di risposta agli incidenti legati a pregiudizi, che obbligano il fornitore a segnalare errori o disuguaglianze individuati che incidono su popolazioni di pazienti differenziate per sesso e genere o per origine razziale/etnica e a delineare le misure correttive;

[38] Commissione europea, *Ora disponibili le clausole contrattuali modello aggiornate dell'UE sull'IA*, Bruxelles: Commissione europea, 2025.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

- (7) chiara attribuzione delle responsabilità, che specifichi gli obblighi del fornitore e del gestore in merito alla conformità ai requisiti dell'AI Act e alla gestione dei rischi legati all'equità e alla parzialità.

Fasi di convalida e pre-implementazione

- **Politica di governance e rappresentatività dei dati sanitari.** Per mitigare i pregiudizi di genere e razziali, tutti gli ospedali e le organizzazioni della società civile (CSO) che implementano l'IA dovrebbero adottare una politica formale che garantisca che i dati utilizzati a fini di addestramento, messa a punto o implementazione siano rappresentativi, completi e accurati per le popolazioni servite. Ciò include la valutazione della composizione demografica, la documentazione delle popolazioni di pazienti sottorappresentate e la definizione di misure di mitigazione ove necessario. La politica dovrebbe riguardare sia i set di dati interni che quelli di provenienza esterna, compresi i dati proprietari e quelli ad accesso aperto, e richiedere la verifica delle prestazioni disaggregate e di eventuali strategie di mitigazione dei pregiudizi applicate dai fornitori. Tale governance è in linea con l'AI Act (art. 10), che richiede l'esame e la mitigazione dei rischi di pregiudizio, e rispecchia le migliori pratiche riportate nella letteratura sulla mitigazione dei pregiudizi nell'IA sanitaria, che copre l'intero ciclo di vita, dalla raccolta dei dati al monitoraggio post-implementazione.[39] Questa politica non è un artefatto in sé, ma un livello di politica che guida altri artefatti. Funziona come principio pre-implementazione e si integra direttamente nel monitoraggio continuo, nell'audit e nella rendicontazione.

- **Linee guida di base sulla sicurezza informatica e modello di minaccia specifico per l'IA.** I sistemi di IA ospedalieri possono essere vulnerabili non solo ad attacchi convenzionali quali l'avvelenamento dei dati, gli esempi avversari o l'evasione del modello, ma anche a manipolazioni o errori che esacerbano i pregiudizi di genere o razziali, ad esempio influenzando in modo sproporzionato gli strati demografici sottorappresentati nei dati di addestramento o di implementazione. L'AI Act (art. 15) richiede resilienza contro le alterazioni non autorizzate dei risultati o delle prestazioni, sottolineando la necessità di misure di salvaguardia specifiche per l'IA. Gli ospedali dovrebbero quindi stabilire una linea di base per la sicurezza informatica e un modello di minaccia specifico per l'IA che consideri esplicitamente i rischi legati ai pregiudizi insieme alle minacce tecniche. Questo documento in continua evoluzione funge sia da linea guida che da strumento pratico, fungendo da lista di controllo pre-implementazione per garantire l'attuazione di controlli di sicurezza informatica allineati con l'MDR e l'IVDR, la governance organizzativa della sicurezza delle informazioni (ad es. ISO/IEC

[39] Char, D.S., Shah, N.H., & Magnus, D., *Implementazione dell'apprendimento automatico nell'assistenza sanitaria — Affrontare le sfide etiche*, *New England Journal of Medicine* 378: 981-983 (2018).

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

27001) e piani di modellizzazione delle minacce e di test sensibili ai pregiudizi.[40] Supporta inoltre il monitoraggio continuo post-implementazione per individuare i rischi emergenti, sia tecnici che legati ai pregiudizi, e per rispondere in modo proattivo. Attraverso l'attuazione di questa politica, gli ospedali possono rafforzare la resilienza dei sistemi, salvaguardare l'equità nei processi decisionali supportati dall'IA e mantenere la conformità con l'AI Act (artt. 9-15) e con i requisiti MDR/IVDR.

- **Un «dossier del sistema di IA» per ogni strumento.** Questo documento è un documento standardizzato redatto per ciascun sistema di IA prima della sua implementazione all'interno di un'organizzazione sanitaria. Il dossier dovrebbe riassumere lo scopo previsto del sistema, l'integrazione nel flusso di lavoro clinico, i risultati della validazione e le metriche di prestazione, con particolare attenzione alle prestazioni nelle diverse popolazioni di pazienti differenziate per sesso e genere o origine razziale/etnica, nonché a eventuali disparità note o misure di mitigazione. Dovrebbe documentare le modalità di supervisione umana, chiarire quali decisioni rimangano esclusivamente sotto il controllo del personale clinico e segnalare eventuali controindicazioni rilevanti per la sicurezza e l'equità dei pazienti. Fornendo una panoramica formalizzata e consapevole dei pregiudizi del sistema di IA, il dossier favorisce la trasparenza, la governance interna e la supervisione normativa, garantendo che le decisioni precedenti all'implementazione tengano esplicitamente conto dei rischi di esiti iniqui.

- **Un «pacchetto combinato di FRIA e DPIA».** Gli ospedali pubblici che implementano sistemi di IA ad alto rischio sono sempre più tenuti a valutare i rischi per i diritti fondamentali, oltre a quelli relativi alla protezione dei dati personali. L'integrazione di queste valutazioni in un unico modello che soddisfi le disposizioni dell'AI Act (art. 27), del GDPR e della DPIA può ridurre l'onere amministrativo, garantendo al contempo che le questioni chiave relative alle popolazioni interessate, ai potenziali danni, alle strategie di mitigazione e ai meccanismi di reclamo siano affrontate in modo sistematico, con particolare attenzione alle potenziali disuguaglianze che interessano le popolazioni di pazienti differenziate per sesso, genere e razza/etnia.

  Il «pacchetto combinato FRIA e DPIA» opera principalmente come strumento di verifica ex ante e aggiornato periodicamente, collocandosi nella fase precedente all'implementazione e all'interno del livello di verifica dell'equità e dei diritti. Queste valutazioni sono analisi organizzate e documentate dei rischi per i diritti

[40] Gruppo di coordinamento sui dispositivi medici, *MDCG 2019-16 Rev. 1 - Linee guida sulla sicurezza informatica per i dispositivi medici*, Bruxelles: Commissione europea, 2020.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

fondamentali e la protezione dei dati, che fungono da punti di controllo della governance piuttosto che da strumenti di monitoraggio continuo. Sebbene debbano essere aggiornate in caso di cambiamenti dei rischi o delle condizioni del sistema, il loro ruolo principale è quello di fornire una valutazione sistematica e verificabile dei potenziali impatti prima e durante l'implementazione, sostenendo la conformità sia alla legge sull'IA (AI Act) che al GDPR e contribuendo a identificare e mitigare i pregiudizi legati al genere o alla razza.

Per rafforzare l'aspetto dei diritti umani, nella progettazione della valutazione d'impatto si dovrebbe prevedere una consultazione significativa con le organizzazioni della società civile (CSO) al fine di integrare la loro competenza in materia di diritti. Ad esempio, la Rete europea delle istituzioni nazionali per i diritti umani (ENNHRI) esorta la Commissione europea a integrare i modelli con linee guida e consultazioni pubbliche per garantire valutazioni d'impatto sui diritti fondamentali (FRIAs) efficaci che tengano conto dell'equità nelle dimensioni di genere e razziali/etniche.[41]

- **Una «politica di controllo delle modifiche».** Tale politica disciplina gli aggiornamenti dei sistemi di IA, compresi il riaddestramento dei modelli, gli aggiornamenti software o le modifiche ai flussi di lavoro clinici. Le organizzazioni dovrebbero definire chiaramente cosa costituisce una «modifica sostanziale» che richiede una nuova convalida o approvazione, assicurando che gli aggiornamenti non introducano rischi imprevisti per l'assistenza ai pazienti. In particolare, le modifiche devono essere valutate in relazione ai potenziali impatti sull'equità di genere e razziale, compresi i cambiamenti nelle prestazioni disaggregate o i risultati differenziali tra popolazioni differenziate in base alle caratteristiche demografiche. Gli ospedali dovrebbero inoltre garantire che i medici e il personale interessato siano informati delle modifiche che potrebbero influire sugli esiti assistenziali per le diverse popolazioni di pazienti, consentendo un'adeguata supervisione e l'adozione di misure di mitigazione. Integrando le considerazioni relative all'equità nella gestione del cambiamento, gli ospedali mantengono sia la sicurezza che l'equità del processo decisionale supportato dall'IA nella pratica clinica.
- **Un «sistema di gestione della qualità» (SGQ) per gli implementatori che sono anche fornitori**. Un'altra dimensione importante del modello riguarda l'IA sviluppata internamente dagli ospedali o dagli istituti di ricerca. Le organizzazioni sanitarie pubbliche realizzano o adattano sempre più spesso i propri strumenti

[41] Rete europea delle istituzioni nazionali per i diritti umani (ENNHRI), *Dichiarazione su come garantire valutazioni d'impatto sui diritti fondamentali (FRIA) efficaci ai sensi della legge dell'UE sull'IA*, Bruxelles: ENNHRI, 2025.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

di IA, in particolare nei centri medici accademici. In tali casi, l'istituzione può di fatto agire come produttore o fornitore, il che comporta ulteriori responsabilità normative. Le linee guida europee sui dispositivi medici sviluppati internamente chiariscono che le istituzioni sanitarie devono gestire sistemi di gestione della qualità e conservare la documentazione tecnica che dimostri che gli strumenti sviluppati internamente rispondono a esigenze specifiche dei pazienti che non possono essere soddisfatte tramite prodotti commerciali.[42] Il SGQ dovrebbe garantire che i risultati dell'IA siano valutati in termini di prestazioni differenziali tra le popolazioni di pazienti differenziate per sesso e genere e per origine razziale/etnica, che le strategie di mitigazione siano documentate e che le modalità di supervisione umana preservino esplicitamente la sicurezza dei pazienti e l'equità. Anche laddove si applichino esenzioni normative per i dispositivi sviluppati internamente, gli obblighi derivanti dai diritti fondamentali e dai quadri normativi più ampi in materia di governance dell'IA rimangono pienamente rilevanti, rendendo il sistema di gestione della qualità uno strumento fondamentale per integrare l'equità nella progettazione, nella convalida e nell'implementazione dei sistemi di IA biomedica.

Fase di monitoraggio, verifica e rendicontazione

- **Un «rapporto sulle prestazioni e sull'equità disaggregato per popolazione».** Poiché le variazioni nelle prestazioni algoritmiche possono contribuire a esiti sanitari diseguali, gli ospedali non solo dovrebbero richiedere ai fornitori di fornire informazioni pertinenti prima dell'appalto, ma dovrebbero anche produrre analisi interne che mostrino come i sistemi si comportano tra pazienti con caratteristiche demografiche diverse. Tali analisi devono concentrarsi sulle prestazioni stratificate per sesso e genere e sulle differenze razziali o etniche rilevanti, ove lecito e fattibile, nonché sulla giustificazione di eventuali divari di prestazione inevitabili. Laddove sussistano divari di prestazioni, le organizzazioni dovrebbero documentare le strategie di mitigazione e monitorare l'efficacia di tali misure nel tempo. Questo approccio riflette le aspettative emergenti nell'ambito della normativa europea sull'IA, secondo cui le organizzazioni devono valutare e mitigare attivamente i rischi di parzialità nei set di dati e nei risultati dei modelli. L'AI Act (art. 10) richiede esplicitamente l'esame e la mitigazione dei rischi di parzialità nei set di dati e consente un trattamento limitato dei dati di categorie speciali per l'individuazione della parzialità, nel rispetto di rigorose garanzie.

[42] Gruppo di coordinamento sui dispositivi medici, *MDCG 2023-1 – Linee guida sull'esenzione per le istituzioni sanitarie (MDR/IVDR)*, Bruxelles: Commissione europea, 2023.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Il “rapporto sulle prestazioni e l’equità disaggregato per popolazione” funge principalmente da strumento di monitoraggio continuo all’interno del quadro di governance. Richiede una raccolta continua di dati, un monitoraggio regolare delle prestazioni del sistema tra i gruppi demografici rilevanti e un aggiornamento costante delle misure di mitigazione laddove vengano individuati risultati differenziali. Ciò consente alle organizzazioni sanitarie di individuare e affrontare i rischi di distorsione nel tempo. Allo stesso tempo, i risultati chiave di questo rapporto, quali le disparità individuate e l’efficacia delle strategie di mitigazione, dovrebbero essere sintetizzati e inclusi nell’«audit sull’equità e i diritti», garantendo trasparenza e responsabilità pubblica.

- **Un regime di audit a due livelli comprendente un audit tecnico e un audit sull’equità e sui diritti.**

  Un regime di audit fattibile per gli ospedali dovrebbe bilanciare il rigore con la praticità ed essere comprensibile sia per il personale che per il pubblico. In questo contesto, un audit è una revisione formalizzata e periodica della conformità, delle prestazioni e degli impatti di un sistema di IA sui diritti, progettata per integrare il monitoraggio continuo piuttosto che sostituirlo.

  Audit tecnico: questo livello costituisce una revisione operativa riservata e comprende controlli dettagliati che rimangono interni, tra cui test di sicurezza, valutazioni di robustezza, convalida delle prestazioni cliniche e conformità alla governance dei dati. Si dovrebbe prestare particolare attenzione alle prestazioni in relazione alle differenze di sesso e genere e alle differenze razziali o etniche, valutando se i risultati del modello svantaggino sistematicamente una qualsiasi popolazione. Queste attività dovrebbero allinearsi alle disposizioni pertinenti dell’AI Act (artt. 9-15), dell’MDR e dell’IVDR, ove applicabili.

  Audit sull’equità e sui diritti: questo secondo livello si concentra sulla trasparenza, la responsabilità e la mitigazione dei pregiudizi. Dovrebbe documentare le disuguaglianze nei risultati, le popolazioni incluse nei test, i limiti noti e i meccanismi di reclamo, con un’analisi esplicita degli impatti di genere e razziali. Dovrebbero inoltre essere raccolte prove che dimostrino che la supervisione umana è significativa e in grado di correggere i risultati distorti. Seguendo l’approccio di «verifica e divulgazione dei pregiudizi» utilizzato nella regolamentazione del lavoro,[43] una sintesi rivolta al grande pubblico può comunicare i risultati chiave, le strategie di mitigazione e i miglioramenti

[43] Dipartimento per la tutela dei consumatori e dei lavoratori della città di New York, *Strumenti automatizzati per le decisioni in materia di occupazione (AEDT)*, New York: Città di New York, 2025.

Finanziato dall’Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell’Unione europea o dell’Agenzia esecutiva europea per l’istruzione e la cultura (EACEA). Né l’Unione europea né l’EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

implementati, senza rivelare informazioni proprietarie o riservate. Anche il «pacchetto combinato FRIA e DPIA» rientra nel livello di verifica dell'equità e dei diritti, rafforzando la valutazione rigorosa dei diritti fondamentali, della protezione dei dati e dell'equità demografica.

- **Un «registro di vigilanza sull'IA».** Il monitoraggio post-implementazione è supportato anche dalla creazione di un registro di vigilanza sull'IA, ispirato ai sistemi di farmacovigilanza utilizzati nella regolamentazione dei medicinali.[44] Gli operatori sanitari dovrebbero poter segnalare incidenti, quasi incidenti o esiti inattesi che coinvolgono decisioni supportate dall'IA, con campi espliciti che rilevino eventuali disparità osservate relative al sesso e al genere o alla razza/etnia. Le segnalazioni dovrebbero essere analizzate sistematicamente per individuare modelli emergenti, compresi gli impatti differenziali su specifici gruppi demografici, consentendo alle organizzazioni di rilevare segnali di sicurezza, preoccupazioni relative all'equità o tendenze di parzialità. Ove necessario, gli incidenti, in particolare quelli che indicano un trattamento iniquo o un rischio per i gruppi protetti, dovrebbero essere comunicati tempestivamente ai fornitori del sistema, ai comitati di governance ospedalieri competenti e alle autorità di regolamentazione. Il registro dovrebbe inoltre alimentare audit periodici sull'equità e sui diritti, a sostegno del miglioramento continuo e della mitigazione della parzialità nell'implementazione clinica dell'IA.

- **Collaborazione a livello di rete e condivisione degli incidenti.** La rete europea di organizzazioni della società civile (CSO) e ospedali pubblici sviluppata nell'ambito del quadro AEQUITAS funge da livello di governance condivisa, fornendo un vocabolario comune e standard di segnalazione per gli incidenti e i risultati relativi ai pregiudizi di genere e razziali nei sistemi di IA. Questa rete sostiene l'apprendimento tra strutture e la diffusione delle migliori pratiche per individuare, mitigare e porre rimedio alle disuguaglianze, analogamente agli approcci adottati nei settori ad alta sicurezza e nella farmacovigilanza. Collegando i dati sugli incidenti di parzialità, sulle prestazioni disaggregate e sulle strategie di mitigazione, la rete consente agli ospedali e alle organizzazioni della società civile di identificare modelli sistemici di disuguaglianza e coordinare interventi mirati. Poiché l'EHDS facilita sia l'uso primario (erogazione delle cure) sia l'uso secondario (ricerca e politiche) in un quadro normativo, la rete può allineare le proprie pratiche di condivisione dei dati e di apprendimento alle tempistiche dell'EHDS e agli atti di esecuzione man mano che questi maturano, sostenendo un'implementazione dell'IA più equa e trasparente tra le istituzioni.

[44] Agenzia europea per i medicinali, *Modulo GVP VI – Raccolta e segnalazione di sospette reazioni avverse (Rev. 2)*, Londra: EMA, 2017.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

- **Modulo di formazione per medici ed esperti delle organizzazioni della società civile.** Per mitigare i pregiudizi legati all'automazione e le disuguaglianze nell'IA clinica, il modello normativo AEQUITAS prevede una formazione obbligatoria volta a fornire agli operatori sanitari le competenze necessarie per comprendere i limiti dei sistemi di IA, interpretare i risultati in modo critico e sovrascrivere o arrestare i sistemi quando i risultati potrebbero produrre esiti distorti o iniqui. La formazione pone l'accento sulla consapevolezza dei rischi di pregiudizi di genere e razziali, sui principi dei diritti fondamentali e sulle considerazioni etiche, tra cui l'equità, la non discriminazione e l'inclusione dei pazienti.

  Questo modulo può essere considerato uno strumento abilitante, che sostiene attivamente l'uso del modello normativo garantendo che il personale responsabile della supervisione dell'IA sia in grado di attuare con competenza le procedure di monitoraggio, convalida e rendicontazione in un'ottica di equità. La formazione prevede esercizi basati su casi concreti relativi all'individuazione e alla mitigazione delle disparità di sesso, genere e razza/etnia.

Conclusione

L'IA offre significative opportunità per migliorare gli esiti sanitari, ma senza un'attenta governance può riprodurre o aggravare le disuguaglianze di sesso e genere e razziali o etniche, nonché altre forme di pregiudizio sistemico. Il modello normativo proposto in questo documento programmatico traduce i principi dei diritti fondamentali in garanzie concrete lungo l'intero ciclo di vita dell'IA, dall'approvvigionamento e dalla convalida all'implementazione, al monitoraggio e alla rendicontazione, con un'attenzione esplicita all'individuazione e alla mitigazione delle disparità negli esiti tra le diverse popolazioni di pazienti. Attraverso l'implementazione di un'architettura di governance dell'IA strutturata attorno a una serie chiara di fasi del ciclo di vita, politiche, strumenti pratici e artefatti, gli ospedali e le organizzazioni della società civile possono valutare, mitigare e documentare sistematicamente i rischi, compresi gli impatti iniqui sulle popolazioni di pazienti emarginate. Integrando le esperienze intersettoriali, i requisiti giuridici dell'UE e i quadri di governance internazionali, il modello normativo supporta gli operatori sanitari nella gestione responsabile dei sistemi di IA e garantisce un'assistenza sanitaria equa, sicura e trasparente per tutti i pazienti, indipendentemente dal sesso e dal genere, dall'origine razziale o etnica o da altre caratteristiche.

Finanziato dall'Unione europea. Le opinioni e i punti di vista espressi sono tuttavia esclusivamente quelli degli autori e non riflettono necessariamente quelli dell'Unione europea o dell'Agenzia esecutiva europea per l'istruzione e la cultura (EACEA). Né l'Unione europea né l'EACEA possono essere ritenute responsabili al riguardo. Codice del progetto: 101215009 — AEQUITAS — CERV-2024-CHAR-LITI

Co-funded by the European Union

AEQUITAS