Università degli studi di Padova

Dipartimento di diritto pubblico, internazionale e comunitario

Corso di Laurea in Diritto e Tecnologia

a.a. 2023/2024

# BIAS NEL TRATTAMENTO DEI DATI MEDICI E SANITARI

Relatrice:
Prof.ssa Annalisa Volpato

Laureanda: Greta Dervishaliaj
Matricola n° 2052106

# BIAS NEL TRATTAMENTO DEI DATI MEDICI E SANITARI

# *INDICE*

*Bibliografia*

*Atti normativi*

*Sitografia*

# *INTRODUZIONE*

L'accesso a un'assistenza sanitaria di qualità è riconosciuto come un diritto fondamentale dall'Unione Europea, sancito dall'Articolo 35 della Carta dei diritti fondamentali dell'Unione Europea, che garantisce il diritto alla protezione della salute e a cure mediche di qualità per i cittadini.[1] Questo principio *fondamentale* riflette l'impegno degli Stati membri a promuovere l'equità e ridurre le disuguaglianze nell'ambito sanitario.
Il progresso tecnologico ha dimostrato di essere un valido alleato della ricerca scientifica e ogni giorno supporta i professionisti sanitari nel loro lavoro. In particolare, i dati sanitari rappresentano una risorsa fondamentale per migliorare la qualità dell'assistenza medica e della ricerca. Tuttavia, il trattamento e analisi di questi dati non sono immuni da distorsioni ed errori sistematici, noti come *bias*, che possono produrre effetti di discriminazione e disuguaglianza nell'accesso a delle cure di qualità.
L'obiettivo di questa tesi è di adottare un approccio multidisciplinare che integri l'analisi tecnica, giuridica ed etica in riguardo al problema dei bias nel trattamento dei dati sanitari.
In primo luogo, si intende esplorare le cause alla radice dei bias nei dati sanitari, con particolare riguardo alla fase di raccolta, trattamento e analisi dei dati.

[1] Art. 35, Carta dei diritti fondamentali dell'Unione Europea

Successivamente, verrà condotta un'analisi giuridica per comprendere se il quadro normativo europeo attuale sia adeguato a prevenire e mitigare i rischi legati ai bias.
Infine, sulla base delle criticità emerse, la tesi propone di individuare possibili strategie per mitigare i rischi legati ai bias, attraverso l'adozione di soluzioni sia tecniche che regolatorie.
L'obiettivo della tesi, comunque, non è solo quello di evidenziare le problematiche esistenti, ma anche di fornire un contributo propositivo per costruire un sistema sanitario che garantisca equità, trasparenza e rispetto dei diritti fondamentali.

# CAPITOLO 1: Bias nel trattamento dei dati sanitari da un punto di vista tecnico

## 1.1 *Bias*: Definizioni e implicazioni

Nell'era moderna, ciò che rappresenta il vero potere e la chiave per lo sviluppo è l'informazione.
La rapida evoluzione tecnologica e la globalizzazione hanno introdotto nuove sfide per la protezione dei dati personali. La raccolta e condivisione di tali dati sono aumentate in maniera significativa, permettendo sia alle imprese private che alle autorità pubbliche di utilizzarli su una scala senza precedenti nello svolgimento delle proprie attività.
Questa trasformazione tecnologica ha modificato profondamente l'economia e le relazioni sociali, facilitando la libera circolazione dei dati personali all'interno dell'Unione e il loro trasferimento verso paesi terzi e organizzazioni internazionali.[2]
Uno dei settori che più beneficia degli effetti positivi della condivisione dei dati è il settore sanitario.
Il trattamento dei dati sanitari rappresenta una componente cruciale per lo sviluppo di politiche sanitarie efficaci, la ricerca medica e la personalizzazione delle cure. Tuttavia, l'analisi e l'interpretazione di questi dati non sono immuni da errori e distorsioni, noti anche come "Bias".
In inglese il Bias è definito come: "an inclination or prejudice for or against one person or group, especially in a way considered to be unfair" e

[2]Considerando 6. REGOLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati)

anche “a systematic distortion of a statistical result due to a factor not allowed for in its derivation”.[3]

Si possono individuare due tipologie di bias: il bias statistico e il bias sociale.

Nella ricerca accademica, il bias statistico si riferisce a un tipo di errore sistematico che può distorcere le misurazioni e/o influenzare le indagini e i loro risultati.[4] È importante distinguere un errore sistematico, come il bias, da un errore casuale. L’errore casuale, infatti, si verifica a causa delle fluttuazioni naturali nell’accuratezza di qualsiasi strumento di misura, delle differenze innate tra gli esseri umani e del puro caso. Gli errori casuali possono verificarsi in qualsiasi momento e sono più difficili da controllare. Gli errori sistematici, invece, si verificano in uno o più punti nel processo di ricerca, come la progettazione dello studio, la raccolta dei dati, l’analisi statistica, l’interpretazione dei risultati e il processo di pubblicazione.[5]

I bias sociali, detti anche bias impliciti, derivano invece da pregiudizi culturali che si riflettono nel modo in cui i dati vengono raccolti e utilizzati. Nello specifico, alcune persone potrebbero essere sottorappresentate nei dataset sanitari, o i modelli di intelligenza artificiale potrebbero perpetuare e amplificare disuguaglianze già esistenti se addestrati su dati che riflettono discriminazioni passate.[6] Il bias implicito si

[3] Oxford English Dictionary, Oxford, OUP, 2023

[4] Popovic A, Huecker MR. Study Bias. 2023 Jun 20. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan–p.1. PMID: 34662027.

[5] Ibidem

[6] https://www.agendadigitale.eu/sanita/ia-emotiva-e-bias-algoritmici-limpatto-nel-settore-sanitario

manifesta in modo automatico e involontario e i risultati più comuni sono di stampo razziale, di genere e di età.[7]

Ad oggi sono numerosi gli esempi discriminatori in ambito di cure e accesso alla sanità nei confronti di alcuni gruppi vulnerabili.

Un esempio di discriminazione razziale nei sistemi di intelligenza artificiale riguarda la diagnosi di cancro alla pelle in pazienti non bianchi. Uno studio del 1979[8] ha dimostrato che la pelle scura possiede melanociti più grandi che producono una maggiore quantità di melanina, proteggendo gli strati più profondi della pelle dagli effetti nocivi del sole.

Una maggiore quantità di melanina epidermica nelle persone di colore filtra almeno il doppio delle radiazioni UV rispetto all'epidermide delle persone bianche. Questo è il motivo per cui la popolazione bianca è la principale vittima del cancro alla pelle e l'incidenza nel cancro è inferiore nelle persone di colore.

Ovviamente, le persone nere possono comunque sviluppare il cancro alla pelle e dei recenti studi hanno dimostrato che sono più propense a morire rispetto ai pazienti bianchi. Infatti, il tasso di sopravvivenza a 5 anni dallo sviluppo del cancro è del 92% per la popolazione bianca e del 70% per la popolazione non bianca.[9]

---

[7] Feng Chen, Liqin Wang, Julie Hong, Jiaqi Jiang, Li Zhou, Unmasking bias in artificial intelligence: a systematic review of bias detection and mitigation strategies in electronic health record-based models, Journal of the American Medical Informatics Association, Volume 31, Issue 5, May 2024, Pages 1172–1183

[8] Kaidbey KH, Agin PP, Sayre RM, Kligman AM. Photoprotection by melanin--a comparison of black and Caucasian skin. J Am Acad Dermatol. 1979 Sep;1(3):249-60. doi: 10.1016/s0190-9622(79)70018-1. PMID: 512075.

[9] Gupta AK, Bharadwaj M, Mehrotra R. Skin Cancer Concerns in People of Color: Risk Factors and Prevention. Asian Pac J Cancer Prev. 2016 Dec 1;17(12):5257-5264. doi: 10.22034/APJCP.2016.17.12.5257. PMID: 28125871; PMCID: PMC5454668.

Uno dei motivi è che nei sistemi di intelligenza artificiale, come le reti neurali convoluzionali (CNN)[10], utilizzate per la classificazione delle lesioni cutanee, mostrano notevoli discrepanze di accuratezza diagnostica tra diversi gruppi etnici. I dataset utilizzati per addestrare queste CNN sono composti principalmente da immagini di lesioni cutanee di pazienti bianchi, con una rappresentazione stimata di pazienti neri solo tra il 5% e il 10%. Infatti, quando testate con immagini di pazienti neri, queste reti presentano un'accuratezza diagnostica circa dimezzata rispetto a quanto dichiarato per i pazienti bianchi.[11]

Per quanto riguarda invece la discriminazione di genere, i dati sulla salute delle donne sono carenti e distribuiti in modo disomogeneo rispetto ad altri dati sanitari generali. Anche quando disponibili, questi dati si concentrano su malattie specifiche, coprendo una limitata gamma di condizioni di salute.

L'esclusione delle donne da numerosi studi clinici crea set dati incompleti, e la mancanza di dati di qualità e rappresentativi sulle condizioni di salute femminili limita l'efficacia di previsione del rischio basati sull'intelligenza artificiale. Questo significa che anche un "buon" algoritmo può presentare forme di discriminazione e sessismo.[12]

[10] Sono un tipo specifico di rete neurale artificiale progettato per il riconoscimento di pattern in dati bidimensionali, come le immagini.

[11] Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. Patterns (N Y). 2021 Oct 8;2(10):100347. doi: 10.1016/j.patter.2021.100347. PMID: 34693373; PMCID: PMC8515002.

[12] Lau, P.L. (2024). AI Gender Biases in Women's Healthcare: Perspectives from the United Kingdom and the European Legal Space. In: Gill-Pedro, E., Moberg, A. (eds) YSEC Yearbook of Socio-Economic Constitutions 2023. YSEC Yearbook of Socio-Economic Constitutions, vol 2023. Springer, Cham.

Uno studio dell'università di Leeds ha dimostrato che dal 2003 al 2013, in UK e Galles, più di 8000 donne sono morte di infarto a causa di un trattamento sanitario ineguale da quelli previsti dagli uomini.
È stato dimostrato che le donne avevano il 34% in meno di probabilità di ricevere un'angiografia coronarica entro 72 ore (24,2% delle donne contro il 36,7% degli uomini), ritardo che comprometteva la diagnosi e il trattamento tempestivo, peggiorando gli esiti clinici. Le donne avevano inoltre il 2,7% in meno di probabilità di ricevere tempestivamente procedure di riperfusioni, il 2,7% in meno di ricevere statine e il 7,4% in meno di ricevere beta-bloccanti alla dimissione (farmaci cruciali per il ridurre il rischio di ulteriori attacchi). Le donne sono il 50% più propense a ricevere una diagnosi iniziale errata rispetto agli uomini, essendo meno probabile che ricevano un elettrocardiogramma pre-ospedaliero, fondamentale per una diagnosi rapida.[13]
Secondo i ricercatori, molte di queste disuguaglianze sono dovuti a bias sistematici che amplificano la concezione errata per cui gli attacchi di cuore siano un problema prevalentemente maschile e dal fatto che gli studi e ricerche al riguardo non tengano adeguatamente in considerazione le differenze biologiche delle donne per permetterne un migliore trattamento.[14]

[13] Wilkinson C, Bebb O, Dondo TB, et al Sex differences in quality indicator attainment for myocardial infarction: a nationwide cohort study Heart 2019;105:516-523.
[14] Ibidem

## *1.2* Bias nel ciclo di vita del trattamento dei medici e sanitari

Esaminiamo ora, da un punto di vista prettamente tecnico, come si genere il bias nel trattamento dei dati sanitari.
Il progresso tecnologico ha causato un aumento significativo di dati generati, ad un livello che è impossibile gestirli con le tecnologie disponibili[15]. Il termine *big data* si riferisce infatti agli "insiemi dei dati raccolti, così vasti e complessi da avere bisogno delle nuove tecnologie, come l'intelligenza artificiale, per venire processati." [16]
Nell'ambito dei big data, sono sei gli attributi associati ai dati, conosciuti anche come le 6 V: Volume, Velocità, Varietà, Valore, Variabilità e Veracità.[17] Applicato in ambito medico, significa che i dati sanitari: (1) hanno un volume importante; (2) crescono a velocità esponenziale; (3) sono generati da fonti diverse; (4) non tutti i dati hanno lo stesso valore e sarà il data scientist di competenza a determinare quali dovrebbero essere conservati e quali eliminati, per variabilità si intende la determinazione della struttura di contestualizzazione del flusso di dati come regolare e affidabile, quindi la capacità di ottenimento di dati rilevanti in tutte le condizioni concepibili. Infine, (5) la veracità si riferisce alla veridicità, accuratezza, affidabilità dell'informazione.[18] *Bias* e anomalie nel trattamento dei dati derivano da un utilizzo di dati non rilevanti o poco

[15] Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, Boccia S. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. Eur J Public Health. 2019 Oct 1;29(Supplement_3):23-27. doi: 10.1093/eurpub/ckz168. PMID: 31738444; PMCID: PMC6859509.
[16] https://www.europarl.europa.eu/topics/it/article/20210211STO97614/big-data-definizione-benefici-e-sfide
[17] https://www.neuralt.com/news-insights/the-6-vs-of-big-data-neural-technologies
[18] G. Manikandan, S. Abirami, K. Gokul, G. Deepalakshmi." Chapter 1: Big data analytics in healthcare theory, tools, techniques and its applications" *Elsevier, 2022*.

accurati per un'analisi. Se non viene fatta una corretta pulizia del dataset in fase preliminare, la veracità si degrada.

Per entrare nello specifico di come si generano i bias da un punto di vista prettamente tecnico, è opportuno analizzare l'intero ciclo di vita del dato e per ogni fase cogliere l'evento o l'assenza dell'evento scatenante.

La prima fase del ciclo di vita del dato è la fase di generazione del dato. In questa fase il dato viene creato dall'interazione tra il paziente e un sistema sanitario. Questo passaggio può sembrare scontato ma perché avvenga la registrazione di un evento clinico, questo deve prima accadere. Se l'evento clinico non accade, il dato non verrà generato e non comparirà quindi in nessun data set. Ci sono diversi fattori che influenzano la registrazione degli eventi clinici e la relativa raccolta dei dati, particolarmente rilevanti per chi analizza dati provenienti da sistemi sanitari diversi. [19]

Per esempio, per quanto riguarda la misurazione della pressione arteriosa (PA), la decisione di effettuare questa procedura dipende dal giudizio del professionista sanitario, tenendo conto delle esigenze mediche del paziente. Tuttavia, anche le differenze nelle politiche nazionali influiscono sulla completezza dei dati relativi alla pressione arteriosa. Nel Regno Unito, gli incentivi e il sistema di rimborso previsto dal paese portano a registrazioni quasi universali della PA[20], mentre nei Paesi Bassi le misurazioni sono più selettive, concentrandosi sui pazienti con malattie

[19] Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. J Med Internet Res. 2018 May 29;20(5):e185. doi: 10.2196/jmir.9134. PMID: 29844010; PMCID: PMC5997930.

[20] https://pharmaceutical-journal.com/article/news/pharmacies-to-get-up-to-1800-in-extra-contract-funding-for-providing-heart-checks

croniche.[21] Importante è rilevare come le politiche nazionali dipendano anche dalle stesse linee guida professionali, che variano tra diversi paesi. Se una linea guida consiglia una misurazione della PA annuale per una determinata popolazione, è più probabile che venga effettuata e registrata. A questo fattore si aggiungono anche le diverse strutture dei diversi sistemi sanitari. Nei sistemi di "gate keeping" il medico di base gestisce l'accesso del paziente agli specialisti, influenzando la raccolta dei dati, ma allo stesso tempo mantengono liste di pazienti stabili, consentendo un monitoraggio più completo nel tempo e creando un denominatore noto per gli studi epidemiologici. A ciò si contrappongono i sistemi non di gate keeping, che non hanno liste fisse di pazienti ma si basano solo su visite saltuarie, limitando la coerenza dei dati e il tracciamento della popolazione. [22]

Infine, un carico di lavoro elevato nelle strutture sanitarie può ridurre la frequenza delle misurazioni regolari della PA.

Ciò significa che gli analisti che utilizzano dati da diversi sistemi sanitari devono essere consapevoli di tali differenze. Per esempio, mediare le registrazioni della PA nel Regno Unito e nei Paesi Bassi usando la popolazione generale come denominatore produrrebbe risultati non validi, poiché nel Regno Unito il sistema sanitario incentiva le misurazioni su un numero maggiore di pazienti rispetto ai Paesi Bassi, dove l'attenzione è rivolta solo a specifici gruppi di pazienti cronici.

---

[21] Carrera PM, Lambooij MS. Implementation of Out-of-Office Blood Pressure Monitoring in the Netherlands: From Clinical Guidelines to Patients' Adoption of Innovation. Medicine (Baltimore). 2015 Oct;94(43):e1813. doi: 10.1097/MD.0000000000001813. PMID: 26512579; PMCID: PMC4985393.

[22] Bartholomeeusen S, Kim CY, Mertens R, Faes C, Buntinx F. The denominator in general practice, a new approach from the Intego database. Fam Pract. 2005 Aug;22(4):442-7. doi: 10.1093/fampra/cmi054. Epub 2005 Jun 17. PMID: 15964863.

Nel caso in cui l'erogazione dell'assistenza avvenga, la seconda fase consiste nella registrazione dell'informazione nei sistemi di cartelle cliniche elettroniche e la rispettiva archiviazione. (EHR).
La registrazione di un evento nelle EHR è cruciale perché i dati siano effettivamente disponibili nei dataset. In questo passaggio gli attori coinvolti sono il professionista sanitario, che esegue la registrazione, e il software alla base delle cartelle elettroniche. [23]
Il primo problema in riguardo a questo è che non tutti i paesi dispongono di una cartella sanitaria elettronica, o comunque non viene utilizzata per qualsiasi operazione, anche se il fenomeno è in crescita soprattutto in seguito alla pandemia del Covid-19 nei paesi OCSE (Organizzazione per la cooperazione e lo sviluppo economico).
Nella fase di archiviazione, nel caso avvenga, in cui i dati vengono salvati per poter essere recuperati, gestiti ed utilizzati anche in futuro, in ambito medico è infatti eseguita attraverso piattaforme elettroniche. Essenziale in questo passaggio è garantire che la copia del dato sia protetta ed accessibile anche nel caso in cui la fonte originale sia danneggiata o compromessa. Le strutture sanitarie devono inoltre adottare misure pratiche e preventive per assicurare la protezione delle informazioni contenute nei loro database.[24] In questo senso sono varie le tecnologie utilizzabili.

[23] Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. J Med Internet Res. 2018 May 29;20(5):e185. doi: 10.2196/jmir.9134. PMID: 29844010; PMCID: PMC5997930
[24] Articolo 32, GDPR

La prima è l'autenticazione. Passwords, sblocco tramite impronta digitale o riconoscimento facciale sono tecniche di identificazione delle persone autorizzate all'accesso dei dati ampiamente utilizzate in ambito medico.
Una seconda tecnica è invece la crittografia dei dati. Si tratta di una tecnica utilizzata quando l'obiettivo principale è quello di garantire sicurezza, riservatezza e integrità del dato in fase di trasmissione o archiviazione del dato. Nella crittografia, l'informazione viene convertita in un formato codificato, detto cifrato. Solo chi dispone della chiave di decodifica può trasformare i dati cifrati nel loro formato originale e leggerli. Alcuni degli algoritmi più utilizzati dalle strutture sanitarie in questo senso sono, RSA, AES, DES, RC4, IDEA e Blowfish.
Infine, un'ulteriore tecnica di garanzia riguarda il controllo accessi. L'utente, controllato da un algoritmo, potrà eseguire solo il compito per cui è autorizzato. Alcuni modelli di controllo accessi popolari per le strutture sanitarie sono il Controllo degli Accessi Basato sui Ruoli e il Controllo degli Accessi Basato sugli Attributi.[25]
Uno studio ha tuttavia rilevato come l'utilizzo di software diversi ha prodotto contraddizioni nei risultati e anche nella capacità di diagnosticare e prescrivere dei farmaci[26]

[25] Ashrafuzzaman, M., Milu, M., Anjum, A., Khanam, F., & Md. A.R. (2022). Chapter 5 - Big data analytics techniques for healthcare. In:Keikhosrokiani, B. P. (ed.) Big Data Analytics for Healthcare. Datasets, Techniques, Life Cycles, Management, and Applications (pp. 49-62). New York: Academic Press.
[26] Van der Bij S, Khan N, Ten Veen P, de Bakker DH, Verheij RA. Improving the quality of EHR recording in primary care: a data quality feedback tool. J Am Med Inform Assoc. 2017 Jan;24(1):81-87

A questo aspetto spesso si aggiunge una mancanza di familiarità con eventuali linee guida che stabiliscono cosa e quando registrare (o la totale assenza di esse) e una mancanza di familiarità con i software in possesso.[27] Altre problematiche in questo passaggio che possono condurre a dati incompleti o non qualitativi sono gli eventuali comportamenti strategici dei professionisti sanitari che potrebbero adattare la registrazione ad incentivi economici, come già visto negli UK, oppure la decisione sempre del personale sanitario del non registrare diagnosi incerte o informazioni sensibili se condivise con altri professionisti.

La terza fase invece consiste nell'estrazione dei dati dai sistemi di cartelle cliniche elettroniche (i dati devono essere estratti per ulteriori analisi e reportistica).

In questo passaggio sono tre gli attori principali che si occupano o influenzano l'estrazione e il trasferimento dei dati.

Innanzitutto, i professionisti sanitari in ruolo di governance che sovrintendono il processo di estrazione, spesso prendendo decisioni su come e quali dati condividere.

I fornitori di software ed esperti di database sono invece responsabili dell'implementazione tecnica dell'estrazione. Tuttavia, strumenti di estrazione diversi e incompatibilità tra sistemi EHR possono causare discrepanze nei dati estratti, risultando in informazioni incomplete o inaccurate. Inoltre, il software di estrazione è spesso protetto da diritti di

[27] Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. Cit.

proprietà intellettuale, il che limita la trasparenza e rende difficile valutare la qualità del processo di estrazione.[28]

Infine, ci sono i pazienti che in base alle normative sulla privacy, in particolare il Regolamento Generale sulla protezione dei dati personali, possono decidere di non consentire alla condivisione dei propri dati rifiutando il consenso o tramite sistemi di *opt-out*.[29]

Il quarto passaggio riguarda la trasformazione dei dati in database per ulteriore analisi e reportistica.

Una volta estratti, i dati devono essere integrati in un database per prepararli ad analisi o reportistica successiva. Gli esperti di database e i tecnici devono gestire la compatibilità dei dati, in particolare quando arrivano in formati e codifiche diversi, che possono cambiare nel tempo.[30] Questo precesso di "integrazione semantica"[31] è essenziale per garantire che i dati siano coerenti e utilizzabili ai fini previsti.

La quinta fase consiste nella preparazione del Dataset per il ricercatore, ovvero la generazione di un file di dati per la ricerca.

In questa fase, per preparare i dati per la ricerca, si selezionano i dati rilevanti sulla base del "need to know", ovvero la domanda di ricerca. La determinazione dei dati necessari per una domanda di ricerca è quindi una responsabilità condivisa tra il ricercatore e il gestore del database, che avranno un ruolo particolarmente significativo sul contenuto del dataset da

[28] Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. Cit.

[29] Art. 21 Regolamento Generale sulla Protezione dei Dati (UE/2016/679)

[30] Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. Cit.

[31] Per integrazione semantica si intende, in generale, all'aggregazione di informazioni da una o più fonti disparate allo scopo di creare un sistema in cui le informazioni sono organizzate in modo sensato per un utente.

analizzare. Questo significa che non tutti i dati che sono stati archiviati verranno effettivamente utilizzati, considerato anche che non tutti gli archivi supportano ogni tipo di ricerca. Nella fase di archiviazione, inoltre, per far fronte alla quantità ingente di dati, comune è anche la tecnica di *data compression*, con cui vengono ridotti i bit necessari per rappresentare il medesimo dato. Le tecniche previste per la compressione dei dati possono essere di tipo: "Lossy compression" (compressione con perdita) oppure "Lossless Compression" (compressione senza perdita).
Quest'ultima permette di comprimere i dati localizzando e rimuovendo eventuali ridondanze statistiche mentre la prima compressa i dati riducendo la complessità delle informazioni oppure con l'eliminazione delle informazioni considerate superflue.  La Lossy compression permette di salvare più spazio rispetto alla Lossless compression; tuttavia, la qualità delle informazioni sarà superiore utilizzando quest'ultima.[32] Il metodo preferito dalla struttura sanitaria di riferimento, quindi, va ad incidere sulla qualità del dato, per non parlare del fatto che regolamenti o comitati direttivi possono comunque negare l'uso degli archivi per scopi specifici, influenzando la completezza dei dati estratti.[33]
La sesta fase è la fase di analisi. Qui comune è la tecnica del *data wrangling* o *data munging*. Si tratta di processo eseguito per rispondere a una domanda specifica, altamente analitica, in cui viene fatta un'analisi

[32]. Moura L, Furuie SS, Gutierrez MA, Tachinardi U, Rebelo MS, Alcocer P, Melo CP. Lossy compression techniques, medical images, and the clinician. MD Comput. 1996 Mar-Apr;13(2):155-9, 172. PMID: 8684278.
[33] Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. Cit.

della qualità del dato attraverso la sua scomposizione, ristrutturazione e arricchimento, in funzione dell'obiettivo preposto.[34]

Apache "Hadoop" è uno dei software più utilizzati in questo senso in ambito sanitario, costituito da strumenti raggruppati per analizzare e processare grandi quantità di dataset.[35] Non si tratta solo di dataset basati sulla sanità, ma anche dei dati generati e ottenuti dai social media e da altri volumi di grandi dimensioni, che possono essere elaborati.

In questa fase il problema, in funzione dei bias statistici, è che l'utilizzo di strumenti diversi per l'analisi spesso porta a risultati differenti pur utilizzando gli stessi identici dati.

Infine, l'ultimo passaggio riguarda la pubblicazione e interpretazione del risultato.

La pubblicazione dei risultati dei modelli analitici creati, dovranno essere sincronizzati con dati già esistenti. La rappresentazione dei risultati deve essere grafica o in un formato tabellare. Tableau, Jupiter, QlikView e Fusioncharts sono tra i software più utilizzati a questo proposito.[36]

Anche qui, l'utilizzo di strumenti di rappresentazione diversi potrebbero portare a interpretazioni diverse sulla base della semplicità o complessità dello strumento utilizzato.

Infine, bisogna ricordare che, per quanto questi strumenti avanzati permettano di raccogliere, analizzare e produrre risultati importanti per la scienza; l'intervento e il controllo umano rimangono essenziali in queste fasi. L'essere umano saprà porre le giuste domande, trovare gli strumenti

[34] https://www.coursera.org/articles/data-wrangling

[35] G. Manikandan, S. Abirami, K. Gokul, G. Deepalakshmi." Chapter 5: Big data analytics in healthcare theory, tools, techniques and its applications" Elsevier, 2022.

[36] Ibidem

più appropriati per lo scopo, affidarsi alle fonti di dati più affidabili. Inoltre, essenziale sempre sarà la sua interpretazione del risultato ma è fondamentale che i ricercatori siano consapevoli delle decisioni prese nei passaggi precedenti e dei potenziali bias accumulati lungo il processo.[37] In particolare, nella fase di interpretazione i ricercatori potrebbero usare test statistici in modo inappropriato o interpretarli in maniera errata, concentrandosi su differenze statistiche che non hanno reale importanza clinica. Oltre a ciò, i ricercatori potrebbero trarre conclusioni causali da studi osservazionali, che non dimostrano una relazione causa-effetto e, infine, potrebbero trasportare i risultati in situazioni o popolazioni che vanno oltre al contesto di studio, rischiando di generalizzare erroneamente.[38]

Infine, studi recenti suggeriscono che nei sistemi di Machine Learning [39] le previsioni fatte dai modelli di apprendimento automatico possono amplificare le distorsioni già presenti nei dati di addestramento. [40]

[37] Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. Cit.

[38] Popovic A, Huecker MR. Study Bias. Cit., p.6

[39] È una sotto branca dell'intelligenza artificiale incentrata sull'utilizzo di dati, detti di addestramento, e algoritmi che permettono ai computer di apprendere da essi, riducendo la necessità di programmazione esplicita.

[40]Melissa Hall, Laurens van der Maaten, Laura Gustafson, Maxwell Jones, Aaron Adcock. A Systematic Study of Bias Amplification. Cornell University, 2022

# Capitolo 2: Quadro giuridico dei dati sanitari e *bias* nel trattamento dei dati sanitari

## 2.1 Raccolta e trattamento dei dati sanitari nel Regolamento 2016/679 UE

Il 25 maggio 2018 è entrata in vigore una delle leggi più importanti, strutturate e rivoluzionarie in materia di dati personali nell'Unione Europea: il Regolamento Generale per la Protezione dei Dati 2016/679, meglio noto come General Data Protection Regulation (GDPR), acronimo che per semplicità verrà utilizzato anche in seguito per far riferimento a suddetto Regolamento.[41]

Il Regolamento mira a creare un equilibrio tra la libera circolazione dei dati e la protezione dei diritti fondamentali delle persone.

Ai fini di questa tesi, è essenziale evidenziare come vi è stata fatta una distinzione tra dati personali e categorie particolari di dati, detti anche, più informalmente, "dati sensibili".

I dati personali sono definiti come: " qualsiasi informazione riguardante una persona fisica identificata o identificabile («interessato»); si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale" [42]

[41] Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati)

[42] GDPR, Articolo 4(1)

Le categorie particolari di dati sono invece definite come tutti quei dati che: "rivelano l'origine razziale o etnica di una persona, le sue opinioni politiche, religiose o filosofiche, l'appartenenza sindacale, nonché i dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute e alla vita od orientamento sessuale della persona."[43]

La decisione di distinguere alcuni dati personali da altri non è una novità del regolamento.

Infatti, già nel 1990, a livello internazionale, le Nazioni Unite nelle "Linee guida per la regolamentazione degli archivi informatizzati di dati personali"[44] considerano i dati sensibili come bisognosi di una protezione ulteriore in quanto potrebbero dar luogo ad un'illecita e arbitraria discriminazione. Similmente, sul piano europeo, quasi 30 anni dopo, il considerando 51 del GDPR evidenzia come "Meritano una specifica protezione i dati personali che, per loro natura, sono particolarmente sensibili sotto il profilo dei diritti e delle libertà fondamentali" e che "il loro trattamento potrebbe creare rischi significativi per i diritti e le libertà fondamentali". Le categorie particolari di dati sono dunque dati il cui trattamento potrebbe comportare rischi relativi a diritti fondamentali e/o libertà personali, come per esempio fenomeni di discriminazione e, come meglio specificato nella relazione esplicativa della Convenzione 108 sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale del Consiglio d'Europa, anche fenomeni portanti a danni

[43] GDPR, Articolo 9

[44] Guidelines for the Regulation of Computerized Personal Data Files, United Nations, 1988

all'integrità fisica o morale di una persona o effetti negativi in riguardo alla presunzione d'innocenza. [45]

La protezione dati di carattere personale viene considerata come diritto fondamentale anche ai sensi dell'articolo 8 della Carta dei Diritti Fondamentali dell'Unione Europea che sancisce il diritto alla loro protezione e trattamento secondo principi di lealtà, finalità determinate, consenso dell'interessato o altro fondamento legittimo[46]; e dall'articolo 16 del trattato sul funzionamento dell'Unione Europea (TFUE).[47]

Rilevante è considerare come la classificazione di un dato come "sensibile" ha sempre portato ad un aperto dibattito tra Stati Membri ed Istituzioni che esprimevano opinioni diverse in merito.[48]

Una classificazione omogenea di un dato come sensibile, dunque, è stata possibile solo grazie all'entrata in vigore del GDPR. La precedente direttiva 95/46/CE[49], infatti, consentiva agli Stati Membri di poter implementare nel proprio ordinamento ulteriori categorie di dati sensibili rispetto a quelle specificate nella stessa direttiva, creando certezza in riguardo a certe categorie di dati ma disomogeneità e problematiche per i titolari del trattamento che desideravano agire in più Stati Membri con normative differenti.[50]

Il GDPR, oltre a negare la possibilità di effettuare queste integrazioni, definisce chiaramente le categorie particolari di dati e ne aggiunge tre

[45] Convenzione sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale Strasburgo, 28 gennaio 1981

[46] Art.8(2), Carta dei diritti fondamentali dell'Unione Europea

[47] Art. 16(1) ,Trattato sul funzionamento dell'Unione Europea

[48] Quinn P, Malgieri G. The Difficulty of Defining Sensitive Data—The Concept of Sensitive Data in the EU Data Protection Framework. German Law Journal. 2021;22(8): p.1587. doi: 10.1017/glj.2021.79

[49] Direttiva 95/46/CE del Parlamento europeo e del Consiglio, del 24 ottobre 1995, relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati

[50] https://protezionedatipersonali.it/regolamento-generale-protezione-dati

rispetto alla direttiva: (i)i dati genetici, (ii) i dati biometrici che permettono una identificazione diretta di una persona fisica e (iii) i dati relativi all'orientamento sessuale della persona. [51]

Il GDPR ha inoltre stabilito delle nuove basi giuridiche per il trattamento di categorie particolari di dati, aggiungendo motivi relativi a finalità di "interesse pubblico sostanziale", "medicina preventiva o del lavoro", "sanità pubblica" e per scopi di "ricerca e/o archiviazione". [52]

Per una protezione a barriera, inoltre, ulteriori requisiti amministrativi sono richiesti per il trattamento di questi dati. Per esempio, nel caso in cui il trattamento di categorie particolari di dati avvenga "su larga scala", vige l'obbligo per il titolare del trattamento di nominare un Responsabile della Protezione dei Dati[53] e di effettuare una Valutazione d'Impatto sulla Protezione dei Dati[54]. Questi requisiti rappresentano un'importante evoluzione rispetto agli obblighi a cui sono generalmente sottoposti il titolare[55] o responsabile[56] del trattamento dei dati personali. In particolare, si può affermare che essi abbiano ampliato il divario che separa i dati sensibili dai dati non sensibili, soprattutto in termini degli oneri potenziali associati al trattamento dei primi.[57]

Nonostante il GDPR sembri chiarire definitivamente la classificazione dei dati come "categorie particolari" bisognose di una maggiore protezione, nuove sfide e problematiche si aprono a causa dell'evoluzione costante della

[51] Art. 9, GDPR
[52] Art. 9(2), GDPR
[53] Art 37(1c), GDPR
[54] Art 37(3b), GDPR
[55] Art.4(7), GDPR
[56] Art.4(8), GDPR
[57] Quinn P, Malgieri G. The Difficulty of Defining Sensitive Data—The Concept of Sensitive Data in the EU Data Protection Framework. German Law Journal. 2021;22(8):1583-1612. doi:10.1017/glj.2021.79

potenza di calcolo e interconnettività online. La creazione di grandi quantità di dati (attraverso IoT, social media, ecc.) intensifica la complessità di determinare quando i dati personali diventano sensibili.

Di fatto, l'aumento dei dataset disponibili e delle capacità di analisi significa che anche dati "innocui" possono essere combinati per trarre conclusioni "sensibili". Questo perché un numero sempre maggiore di campioni di dati può essere considerato un dato personale, cioè informazioni che possono essere collegate, direttamente o indirettamente, a individui specifici e la natura mutevole dei dati personali significa che potrebbe non essere molto difficile per un titolare del trattamento utilizzare i dataset in suo possesso per arrivare a conclusioni che potrebbero essere di natura sensibile.[58]

Entrando finalmente nel merito dei dati relativi alla salute, questo fenomeno è sempre più in crescita e di particolare rilievo.

I dati relativi alla salute sono definiti dal GDPR come "I dati personali attinenti 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"[59]. Definizione che può potenzialmente includere un'ampia categoria di dati, in base all'accuratezza e alla "quantità" di informazione rivelata sulla persona fisica in questione.

Il considerando 35 del GDPR specifica che "Nei dati personali relativi alla salute dovrebbero rientrare tutti i dati riguardanti lo stato di salute dell'interessato che rivelino informazioni connesse allo stato di salute fisica o mentale passata, presente o futura dello stesso."

[58] Malgieri, G., & Comandé, G. (2017). Sensitive-by-distance: quasi-health data in the algorithmic era. Information & Communications Technology Law, 26(3), 229–249

[59] Articolo 4(15), GDPR

Anche il Gruppo di Lavoro articolo 29 si esprime sostenendo che un dato è considerato relativo alla salute se è intrinsecamente legato alla salute stessa o se, anche se grezzo, può essere combinato con altri dati per trarre conclusioni riguardanti lo stato di salute di una persona e/o i rischi associati[60].

In particolare, vengono individuati due importanti elementi per la definizione di un dato come di natura sensibile.[61]

Il primo elemento è la sensibilità intrinseca di un particolare dataset, ovvero "quanto" sensibile sia l'informazione contenuta nel dato. Il secondo elemento è la "distanza computazionale", ovvero lo sforzo necessario a trarre conclusioni sensibili partendo da dati che a prima vista non appaiono sensibili.

Per esempio, una diagnosi o una certificazione sportiva possono essere considerate informazioni intrinsecamente sensibili, in quanto rivelano in modo diretto e non equivoco informazioni sullo stato di salute di un soggetto.

Informazioni invece relative alle abitudini alimentari, all'esercizio fisico o alla qualità dell'aria di una determinata città, non possono definirsi "intrinsecamente" sensibili, tuttavia, la distanza computazionale tra questi dati e informazioni sullo stato di salute di una persona può essere relativamente corta e fornire informazioni intime ai titolari del trattamento.

[60] Documento di lavoro sul trattamento dei dati personali relativi alla salute contenuti nelle cartelle cliniche elettroniche (CCE), Gruppo di lavoro articolo 29, 15 febbraio 2007

[61] Vayena, E., Gasser, U. (2016). "Strictly Biomedical? Sketching the Ethics of the Big Data Ecosystem in Biomedicine". In: Mittelstadt, B., Floridi, L. (eds) The Ethics of Biomedical Big Data. Law, Governance and Technology Series, vol 29. Springer, Cham.

Un esempio interessante riguarda la distanza computazionale tra dati relativi alle ore passate sui social media e l'identificazione di sintomi depressivi sullo stesso soggetto. Uno studio[62] ha dimostrato che elementi estratti dai dati dei sensori dei telefoni, come il GPS e le ore passate sul telefono, hanno fornito indicatori comportamentali strettamente associati alla gravità dei sintomi depressivi. Sebbene i partecipanti non avessero sintomi clinici confermati, lo studio evidenzia come i dati dei telefoni cellulari potrebbero essere utilizzati per monitorare le popolazioni e mettere quindi a rischio le libertà fondamentali delle persone.

Questo studio mette inoltre in luce come il mondo dell'online, contraddistinto da una forte interconnessione, e il continuo sviluppo della natura dei dati personali e sensibili, permette ai titolari del trattamento di avere una maggiore disponibilità di accesso ad informazioni rientranti nelle categorie particolari di dati, senza però adempiere alla normativa e i requisiti stringenti del GDPR in merito al trattamento di tali dati.

È quindi importante chiedersi se i quadri normativi sulla protezione dei dati, come il GDPR, siano in grado di prevenire o regolamentare i tipi di danno tradizionalmente associati ad uno scorretto trattamento dei dati sensibili, ad esempio i danni associati a contesti discriminatori.[63]

[62] Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, Mohr DC. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res. 2015 Jul 15

[63] Quinn P, Malgieri G. The Difficulty of Defining Sensitive Data—The Concept of Sensitive Data in the EU Data Protection Framework. Cit.

## 2.2 Bias nel trattamento dei dati: normativa di riferimento

I *bias* nel trattamento dei dati medici e sanitari non hanno una normativa specifica di riferimento e la loro regolamentazione va individuata in atti legislativi differenti.

Innanzitutto, considerata la natura discriminatoria del bias, il cui problema principale si riferisce ad una sottorappresentazione delle minoranze etniche e di genere all'interno della popolazione, il primo diritto a cui ci si può appellare è il diritto a non essere discriminati.

Il diritto a non essere discriminati è sancito, nell'Unione Europea, dall'articolo 21 dalla Carta dei Diritti Fondamentali dell'Unione Europea (2000), che sancisce il divieto di esercitare qualsiasi forma di discriminazione fondata sul sesso, la razza, il colore della pelle, l'origine etnica o sociale, le caratteristiche genetiche, la lingua, la religione, le convinzioni personali, le opinioni politiche o di qualsiasi altra natura, l'appartenenza ad una minoranza nazionale, il patrimonio, la nascita, la disabilità, l'età o l'orientamento sessuale.[64]

In particolare, l'Unione Europea proibisce due forme di discriminazione: la discriminazione diretta e la discriminazione indiretta.

"Si ha discriminazione diretta quando una persona è trattata meno favorevolmente di quanto sia, sia stata o sarebbe trattata un'altra (persona) in una situazione analoga in ragione del fattore protetto (ad esempio l'etnia, il sesso o l'orientamento sessuale); la discriminazione è invece indiretta quando la disposizione o l'atto discriminatorio siano apparentemente neutri ma possano mettere la persona portatrice del fattore

[64] Art. 21(1) Carta dei diritti fondamentali dell'Unione Europea

protetto in una situazione di particolare svantaggio rispetto ad altre persone".[65]

Secondo il diritto europeo, sia che ci sia una discriminazione diretta che ci sia una discriminazione indiretta, è irrilevante che l'organizzazione discrimini intenzionalmente o per errore.

Nel caso di decisioni prese tramite IA, le leggi antidiscriminazione possono essere utilizzate per contrastare decisioni discriminatorie del sistema. Tuttavia, queste leggi presentano diverse debolezze, e soprattutto nel caso di discriminazione indiretta dell'IA, non vengono fornite regole chiare, specifiche e settoriali per rispondere al problema.[66]

### 2.2.1 *Bias* e GDPR

Un'altra fonte normativa rilevante nell'ambito dei *bias* è lo stesso GDPR. Innanzitutto, si può rilevare come buona parte dei fattori protetti (Ad esempio etnia e orientamento sessuale) dalla normativa europea antidiscriminazione coincidano con le categorie particolari di dati previste dall'articolo 9 del GDPR. Questo evidenzia come evitare che ci siano forme di discriminazione nel trattamento dei dati sia un obiettivo implicito della creazione di queste categorie particolari di dati.[67]

Un'altra menzione alla discriminazione viene fatta nell'articolo 22, nel quale si stabilisce che l'interessato ha il diritto a non essere sottoposto ad

[65] Dott.ssa Francesca Spena, approfondimento presentato nell'incontro di studio T19010, "Inadempimento, illecito e risarcimento del danno nel rapporto di lavoro – Discriminazione: la definizione nel solco del principio di tipicità", pp 6-7.

[66] Borgesius, Frederik Zuiderveen. Study on Discrimination, Artificial Intelligence, and Algorithmic Decision-Making. Council of Europe, Directorate General of Democracy, 2018. Pp 33-35

[67] Marvin van Bekkum, Frederik Zuiderveen Borgesius. Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception. "Computer Law and Security Review" V.48. p.105-770. 2023.

una decisione basata su un trattamento automatizzato, compresa la profilazione che produca effetti giuridici o incida in modo analogo significativamente sulla sua persona.[68] Ovviamente, questi trattamenti sono comunque consentiti se espressamente previsto dal diritto dell'Unione Europea, ma, e qui ci si addentra in modo significativo in qualcosa di rilevante per i bias:" il titolare del trattamento attua misure appropriate per tutelare i diritti, le libertà e i legittimi interessi dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione."[69] Nello specifico delle categorie particolari di dati, tra cui i dati medici e sanitari, le decisioni prese esclusivamente sulla base del trattamento automatizzato sono sempre vietate a meno che ci sia il (i)consenso esplicito dell'interessato[70], (ii)ci siano motivi di interesse pubblico rilevante[71] e (iii) siano in vigore misure adeguate a tutela dei diritti, delle libertà e legittimi interessi dell'interessato.[72]
L'articolo 22 del GDPR è quindi cruciale per affrontare i bias nei trattamenti automatizzati e di profilazione, poiché responsabilizza il titolare del trattamento a implementare misure tecniche e organizzative adeguate a evitare inesattezze nei dati[73], e di conseguenza, come abbiamo visto nel precedente capitolo, evitare una possibile fonte di bias con tutte le conseguenze che ne derivano.

[68] Art. 22(1), GDPR
[69] Art. 22(3), GDPR
[70] Art.9 (2a), GDPR
[71] Art.9 (2g), GDPR
[72] Art.22(4), GDPR
[73] Considerando 71, GDPR

Chiaramente, l'articolo 22 ha un ambito di applicazione limitato ai processi automatizzati e alla profilazione, quindi il GDPR non sembra fornire particolare tutela nei confronti delle possibili discriminazioni derivanti dai bias.

Per questo motivo, nella giurisprudenza europea si è aperto un dibattito sulla necessità di adottare una nuova eccezione, rispetto a quelle già previste dal GDPR[74], per l'utilizzo delle categorie particolari di dati ai fini di combattere la discriminazione. Questa misura è già stata adottata da sei paesi extra-ue (Bahrain, Curacao, Ghana, Jersey, Sint Maarten e Sudafrica)[75]

Gli argomenti a favore dell'adozione di un'eccezione che consenta l'uso di categorie particolari di dati per prevenire la discriminazione nei sistemi di IA sono i seguenti. Innanzitutto, questa eccezione avrebbe un'utilità pratica per un'autovalutazione del rischio di discriminazione interna. Infatti, se le organizzazioni potessero raccogliere ed esaminare i dati "sensibili", potrebbero verificare più agevolmente se i propri sistemi di IA o algoritmi stiano discriminando in base ai fattori protetti.[76]

Il secondo punto a favore riguarderebbe invece la possibilità per le organizzazioni di dimostrare il rispetto delle leggi antidiscriminazioni con una maggiore facilità. Il monitoraggio dell'equità dei propri sistemi, infatti, sarebbe ideale per rispondere agli obblighi legali di non-

[74] Articolo 9 GDPR

[75] Articolo 5 del Bahrain Personal Data Protection Act 2018; Capitolo 2(2) Curacao National Ordinance Protection of personal data; Ghana Data Protection Act 2012; Jersey Data Protection Act 2012; South Africa Protection of Personal Information Act 2013; Saint Marrin Personal Data Protection Act 2010.

[76] Marvin van Bekkum, Frederik Zuiderveen Borgesius. Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception?, Computer Law & Security Review,Volume 48,2023,105770,ISSN 0267-3649,

discriminazione. A questo si collegano dei benefici per le autorità pubbliche o gli enti per l'uguaglianza, che potrebbero svolgere un ruolo maggiormente attivo nel garantire e monitorare che le organizzazioni rispettino gli obblighi di non discriminazione.[77]

L'ultimo punto a favore invece riguarderebbe la funzione simbolica dell'introdurre una nuova eccezione per far fronte alla discriminazione. Infatti, per le aziende/organizzazioni mostrare un impegno nel contrastare la discriminazione e attenzione agli effetti sociali della tecnologia, darebbe un segnale positivo all'opinione pubblica e potenziali clienti (in particolare le vittime potenziali di queste discriminazioni), incrementando significativamente la sua reputazione.[78]

Gli argomenti che invece vanno contro all'introduzione di un'eccezione che consenta l'uso di categorie particolari di dati per prevenire la discriminazione, evidenziano rischi significativi per la privacy, la sicurezza dei dati e la possibilità di abusi.[79]

In particolare, se le organizzazioni sono autorizzate a raccogliere dati sensibili per finalità di controllo ed equità, si potrebbe verificare un aumento della raccolta dei dati su gruppi già vulnerabili, rendendole ancora più esposte a dinamiche di sorveglianza e discriminazione, producendo quindi effetti opposti a quelli sperati.[80]

Un altro punto a sfavore è relativo al fatto che la mera immagazzinazione dei dati sensibili può rappresentare un'intrusione della privacy, come

[77] Ibidem
[78] Ibidem
[79] Marvin van Bekkum, Frederik Zuiderveen Borgesius. Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception.Cit pp 9-10.
[80] Ibidem

riconosciuto dalla Corte di Giustizia dell'Unione Europea e della Corte dei Diritti dell'uomo, aumentando così i "danni recepiti" da parte delle persone cui i dati sono trattati.[81]

Un altro punto a sfavore riguarderebbe il rischio di abuso e uso improprio dei dati, infatti, ogni raccolta di dati sensibili comporta un rischio intrinseco di violazione della sicurezza. Un eventuale *data breach* potrebbe esporre dati estremamente sensibili, come etnia o preferenze religiose, causando gravi danni agli individui colpiti. Inoltre, la creazione di un'eccezione potrebbe aprire la porta a usi futuri non previsti e potenzialmente pericolosi. [82]

In conclusione, l'introduzione di una nuova eccezione al GDPR per l'uso di dati sensibili, al momento presenta ancora dubbi sulla possibilità che i benefici superino i potenziali rischi del momento. Nel caso in cui questa eccezione venga adottata in futuro, dovranno essere previsti rigorosi meccanismi di sicurezza per la minimizzazione di questi rischi.[83]

### 2.2.2 Bias e AI ACT

Un'altra fonte particolarmente rilevante per far fronte ai bias nel trattamento dei dati sanitari si può individuare nel "AI Act"[84], ovvero il regolamento sull'intelligenza artificiale entrato in vigore il primo agosto

[81] Ibidem
[82] Ibidem
[83] Ibidem
[84] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale)

2024, dopo la sua pubblicazione nella Gazzetta Ufficiale dell'Unione Europea il 12 luglio 2024  per la promozione di un uso sicuro e responsabile dell'intelligenza artificiale, garantire il rispetto dei diritti fondamentali, sostenere l'innovazione e potenziare la competitività dell'UE nel settore .[85]

In particolare, l'AI Act adotta una tecnica di classificazione dei sistemi di intelligenza artificiale in diverse categorie di rischio, stabilendo obblighi per i fornitori e distributori in rapporto alla categoria di rischio. Questa classificazione è determinata dall'uso e dall'impatto potenziale dell'IA sulla salute, sulla sicurezza e sui diritti fondamentali delle persone.[86]

Il più alto grado di rischio è il rischio "inaccettabile". Le pratiche rientranti in questa categoria sono vietate.[87] Nel caso della sanità, proibita è la pratica del "social scoring"[88] basata su sistemi di sorveglianza biometrica per beneficiare dei servizi sanitari. Per i fornitori[89]" e "utenti"[90], sarà quindi vietata l'immissione nel mercato di questi sistemi.[91]

In ambito sanitario, sono considerati sistemi ad alto rischio nell'AI Act, tutti i dispositivi medici superiori alla classe I del Regolamento relativo a

[85] https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_it
[86] Articolo 9 IA Act
[87] Articolo 5 IA aCT
[88] Il "social scoring" sanitario è un sistema che valuta il comportamento e lo stile di vita di una persona, basandosi su dati sanitari e abitudini, per incentivare pratiche salutari e ridurre i costi. Può influire su premi assicurativi e accesso ai servizi, ma solleva questioni etiche legate alla privacy e alla discriminazione.
[89] Una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che sviluppa un sistema di IA o che fa sviluppare un sistema di IA al fine di immetterlo sul mercato o metterlo in servizio con il proprio nome o marchio, a titolo oneroso o gratuito (Art. 3(2), AI Act).
[90] Qualsiasi persona fisica o giuridica, autorità pubblica, agenzia o altro organismo che utilizza un sistema di IA sotto la sua autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale (Art. 3(4), AI Act
[91] Articolo 5, AI Act

dispositivi medici[92] e tutti i dispostivi rientranti nella classe A del Regolamento relativo ai dispositivi medici in vitro[93]. Tra questi ci sono: l'IA per la valutazione del rischio e la determinazione dei prezzi per l'assicurazione sanitaria; l'IA per la classificazione delle chiamate di emergenza e quindi per decisioni relative all'invio di soccorsi medici; l'IA per sistemi di "triage"[94] per i pazienti di emergenza sanitaria e, infine, l'IA utilizzata dalle autorità pubbliche per valutare l'idoneità a ricevere benefici e servizi di assistenza pubblica essenziali, compresi i servizi sanitari.[95] Questa classificazione è piuttosto rilevante per i bias nel trattamento dei dati medici e sanitari perché vengono imposti requisiti rigorosi nei confronti dei fornitori e utenti per garantire che i dati siano di alta qualità, rappresentativi e privi di pregiudizi che potrebbero svantaggiare alcuni gruppi demografici. [96]

Il primo articolo di particolare importanza è l'articolo 10, riguardante la qualità e governance dei dati. È uno degli articoli chiave per la gestione dei dati nei sistemi di intelligenza artificiale ad alto rischio, come quelli utilizzati nel settore sanitario. L'articolo 10 stabilisce che i sistemi IA devono essere addestrati, validati e testati con dati che siano accurati, completi, rappresentativi e pertinenti rispetto allo scopo per cui vengono

[92] Regolamento (UE) 2017/745 del Parlamento europeo e del Consiglio, del 5 aprile 2017, relativo ai dispositivi medici, che modifica la direttiva 2001/83/CE, il regolamento (CE) n. 178/2002 e il regolamento (CE) n. 1223/2009 e che abroga le direttive 90/385/CEE e 93/42/CEE del Consiglio (Testo rilevante ai fini del SEE.

[93] Regolamento (UE) 2017/746 del Parlamento europeo e del Consiglio, del 5 aprile 2017, relativo ai dispositivi medico-diagnostici in vitro e che abroga la direttiva 98/79/CE e la decisione 2010/227/UE della Commissione (Testo rilevante ai fini del SEE. )

[94] Il "Triage" è un sistema utilizzato per selezionare i soggetti coinvolti in infortuni secondo classi di urgenza/emergenza crescenti, in base alla gravità delle lesioni riportate e del loro quadro clinico.

[95] The EU Artificial Intelligence Act (2024): Implications for healthcare

[96] Art. 10, IA Act

utilizzati. Elementi fondamentali, come abbiamo visto, per garantire dataset accurati e con ridotto contenuto di bias.

In dettaglio, l'Articolo 10 richiede che i set di dati di addestramento[97], convalida[98] e prova[99], nell'ambito dei sistemi di IA ad alto rischio, devono essere gestiti tramite pratiche di governance e gestione adeguate alla finalità prevista dal sistema. Alcuni di questi requisiti sono: (1)la definizione di processi precisi per la raccolta dei dati personali, considerando anche la finalità originaria se si tratta di dati personali, per garantire trasparenza e rispetto della privacy; (2)il trattamento dei dati deve includere operazioni come annotazione, etichettatura, pulizia, aggiornamento, arricchimento e aggregazione, assicurando che i dati siano idonei all'uso nel sistema IA;(3) devono essere esplicitate le ipotesi che determinano come i dati vengono interpretati e rappresentati, garantendo che queste siano adeguate alla misurazione e alla rappresentazione prevista; (4)deve essere fatta una valutazione della disponibilità, quantità e adeguatezza dei data set utilizzati; (5) deve essere fatto un esame atto a valutare le possibili distorsioni (quindi bias) suscettibili di incidere sulla salute e la sicurezza delle persone, e di avere un impatto negativo sui diritti fondamentali o di comportare discriminazioni vietate dal diritto dell'Unione, in particolare laddove gli output[100] di dati influenzano gli

[97] I dati utilizzati per addestrare un sistema di IA adattandone i parametri che può apprendere (Articolo 3, IA Act)
[98] I dati utilizzati per fornire una valutazione del sistema di IA addestrato e per metterne a punto, tra l'altro, i parametri che non può apprendere e il processo di apprendimento, al fine tra l'altro di evitare lo scarso (underfitting) o l'eccessivo (overfitting) adattamento ai dati di addestramento (Articolo 3, IA Act)
[99] I dati utilizzati per fornire una valutazione indipendente del sistema di IA al fine di confermarne le prestazioni attese prima della sua immissione sul mercato o messa in servizio (Articolo 3, IA Act)
[100] "Nelle elaborazioni elettroniche, dati che costituiscono il risultato finale dell'elaborazione" Dizionario Treccani

input[101] per operazioni future. Si collega a ciò l'obbligo di adottare misure adeguate alla prevenzione e attenuazione delle possibili distorsioni. Inoltre, i set di dati devono tenere conto delle caratteristiche o degli elementi particolari dello specifico ambito geografico contestuale, comportamentale o funzionale del quale il sistema ad alto rischio è destinato a essere usato, solo nella misura in cui è necessario per la finalità prevista.[102]

Gli obblighi normativi sopracitati riguarderebbero però i dataset composti da dati personali "generici". Tuttavia, il comma 5 dell'articolo 10 stabilisce che i fornitori di sistemi di IA ad alto rischio possono eccezionalmente trattare categorie particolari di dati personali solo quando strettamente necessario per rilevare e correggere distorsioni, come in un punto precedente dell'articolo[103]. Tuttavia, questo trattamento deve avvenire con adeguate tutele per i diritti e le libertà fondamentali delle persone, in conformità ai regolamenti (UE) 2016/679[104] e (UE) 2018/1725[105] e alla direttiva (UE) 2016/680[106], e solo se vengono rispettate tutte le seguenti condizioni: (1)il rilevamento e la correzione efficace delle distorsioni non è possibile con il trattamento di altri dati, inclusi quelli

[101] In informatica, insieme dei dati di ingresso forniti dall'utente al calcolatore" Dizionario Treccani
[102] V. in analogia principio di minimizzazione dei dati, Art. 5 GDPR
[103] Paragrafo 2 punti g e h
[104] Regolamento di esecuzione (UE) 2018/679 della Commissione, del 3 maggio 2018, che rinnova l'approvazione della sostanza attiva forchlorfenuron in conformità al regolamento (CE) n. 1107/2009 del Parlamento europeo e del Consiglio relativo all'immissione sul mercato dei prodotti fitosanitari e che modifica l'allegato del regolamento di esecuzione (UE) n. 540/2011 della Commissione
[105] Regolamento (UE) 2018/1725 del Parlamento europeo e del Consiglio, del 23 ottobre 2018, sulla tutela delle persone fisiche in relazione al trattamento dei dati personali da parte delle istituzioni, degli organi e degli organismi dell'Unione e sulla libera circolazione di tali dati, e che abroga il regolamento (CE) n. 45/2001 e la decisione n. 1247/2002/CE
[106] DIRETTIVA (UE) 2016/680 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 27 aprile 2016relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati e che abroga la decisione quadro 2008/977/GAI del Consiglio

sintetici o anonimizzati; (2)le categorie particolari di dati sono soggette a limitazioni tecniche relative al riutilizzo del dato, nonché a misure di sicurezza avanzate, come la pseudonimizzazione; (3)le categorie particolari di dati sono soggette a misure aggiuntive che garantiscono che i dati personali siano sicuri e accessibili solo a persone autorizzate vincolate da obblighi di riservatezza, con un controllo rigoroso degli accessi e della documentazione per prevenire gli abusi; (4)che i dati non vengano trasmessi, trasferiti o resi accessibili a terzi;(5) che i dati vengano cancellati non appena la distorsione è stata corretta o al termine del loro periodo di conservazione, a seconda dell'evento che si verifica prima e, infine,(6) che compaiano nei registri delle attività di trattamento, come previsto dai regolamenti (UE) 2016/679 e (UE) 2018/1725 e dalla direttiva (UE) 2016/680, le motivazioni per cui il trattamento di queste categorie di dati è stato ritenuto strettamente necessario e perché non è stato possibile utilizzare altri dati a questo scopo.

Questo articolo è particolarmente importante anche perché costituirebbe un importante garanzia per una possibile nuova eccezione, di cui discusso prima[107], con cui sarebbe possibile trattare le categorie particolari di dati se questo trattamento fosse strettamente legato alla prevenzione della discriminazione generata dall'intelligenza artificiale. L'articolo 10 garantisce che vengano adottate salvaguardie appropriate sia da un punto di vista tecnico, con tecniche di anonimizzazione e crittografia, che da un punto di vista giuridico, rispettando i principi del GDPR e garantendo che i dati siano trattati in modo proporzionato e trasparente.[108]

[107] Capitolo 2.2.1
[108] Capitolo 2.2.1

L’articolo 13 del IA Act invece, riguarda gli obblighi di trasparenza e informazione agli *utenti*.

Ai sensi di questo articolo, i sistemi di IA ad alto rischio devono essere sviluppati in modo tale da garantire adeguata trasparenza, consentendo agli utenti di comprendere l’output del sistema e utilizzarlo in linea al suo scopo. In questo senso, ogni sistema deve essere accompagnato da istruzioni dettagliate, disponibili in formato digitale o non digitale che siano coincise, complete, chiare, corrette, accessibili e comprensibili agli utenti. Questo articolo vuole far fronte ai gravi problemi di trasparenza, tipici dell’intelligenza artificiale, che per la loro complessità tecnica e opacità intrinseca rende difficile comprendere il loro funzionamento. La mancanza di trasparenza e salvaguardia istituzionale può provocare gravi conseguenze ai diritti fondamentali, impedendo alle persone o gruppi di persone offese di esercitare i propri diritti.[109]

La trasparenza dell’intelligenza artificiale è fondamentale per consentire un’analisi critica e sistematica dei potenziali errori o limitazione nei dati e nei modelli, migliorando l’equità e l’affidabilità degli algoritmi.[110] In particolare, in ambito clinico, una maggiore trasparenza dei dati utilizzati permetterebbe ai ricercatori di capire se i modelli sono applicabili a scenari clinici reali e in particolare stabilire se l’algoritmo funzioni in contesti diversi o per popolazioni varie.[111]

[109] Busuioc M, Curtin D, Almada M. Reclaiming transparency: contesting the logics of secrecy within the AI Act. European Law Open. 2023;2(1):79-105. doi:10.1017/elo.2022.47

[110] Daneshjou, R., Smith, M. P., Sun, M. D., Rotemberg, V., & Zou, J. (2021). Lack of Transparency and Potential Bias in Artificial Intelligence Data Sets and Algorithms: A Scoping Review. JAMA Dermatology, 157(11), 1362–1369

[111] Ibidem

Una maggiore trasparenza permetterebbe inoltre all'essere umano di poter identificare eventuali bias, come la mancanza di rappresentatività demografica nel dataset. A ciò si lega il fatto che algoritmi allenati su dati non trasparenti possono perpetuare pregiudizi esistenti, mentre con dati trasparenti e accessibili è possibile esaminare e correggere questi pregiudizi.[112]

L'articolo 14 dell'IA Act richiede poi un efficace supervisione da parte di persone fisiche durante il periodo in cui il sistema ad alto rischio è in uso. La sorveglianza umana ha l'obiettivo di prevenire o ridurre al minimo i rischi per la salute, la sicurezza e i diritti fondamentali. La persona o le persone fisiche di sorveglianza dovranno, tra le altre, restare consapevoli della possibile tendenza a fare "automaticamente affidamento o a fare eccessivo affidamento sull'output prodotto da un sistema di IA ad alto rischio («distorsione dell'automazione»), in particolare in relazione ai sistemi di IA ad alto rischio utilizzati per fornire informazioni o raccomandazioni per le decisioni che devono essere prese da persone fisiche"[113]. Inoltre, potranno decidere in qualsiasi momento di ignorare, annullare o ribaltare l'output del sistema AI ad alto rischio.

Questo articolo vuole quindi garantire una protezione maggiore, permettendo ad un soggetto umano di verificare, controllare e gestire l'output di un sistema che sulla base dell'articolo precedente dovrebbe essere chiaro e trasparente. Per quanto possa tornare utile per l'identificazione di bias e rischi per le libertà fondamentali, ricordiamo che

[112] Ibidem
[113] IA ACT, Articolo 14(4b)

anche gli esseri umani sbagliano e sono soggetti a bias sociali. Il che significa che la tutela in tema di bias non è completamente garantita.[114]
Infine, l'articolo 15 si riferisce all'accuratezza, robustezza e cibersicurezza dei sistemi ad alto rischio. Questi requisiti mirano a garantire il funzionamento sicuro e affidabile dei sistemi IA durante il loro intero ciclo di vita, cercando di minimizzare rischi e vulnerabilità.
I sistemi dovranno essere, infatti, quanto più resilienti possibile per quanto riguarda errori, guasti o incongruenze che possono verificarsi nel sistema o nell'ambiente in cui esso opera. Dovranno inoltre essere sviluppati in modo tale di eliminare o ridurre il più possibile il rischio di output distorti che influenzano input di operazioni future (feedback loops).[115]
Infine, per quanto riguarda la cibersicurezza, i sistemi dovranno essere progettati in modo tale da prevenire attacchi che cercano di manipolare il set di dati di addestramento (*data poisoning*) o dei modelli pre-addestrati (*model poisoning*) o degli input progettati per far si che il sistema commetta un errore (*model evasion*).[116]

[114] https://policy-lab.ec.europa.eu/news/fair-decision-making-can-humans-save-us-biased-ai-2024-03-22_en
[115] IA ACT, Articolo 15(4b)
[116] IA ACT, Articolo 15(5)

# CAPITOLO 3: Possibili metodi tecnici e giuridici per il superamento dei bias nel trattamento dei dati medici e sanitari

## 3.1 Utilizzo di dati sintetici per il superamento dei bias

I dati sintetici sono una tipologia di dato generato artificialmente utilizzando dati originali e modelli addestrati per replicare le caratteristiche e la struttura dei dati di partenza. L'obiettivo del loro utilizzo è che producano risultati simili a quelli prodotti dai dati originali in fase di analisi statistica.

Il processo di generazione, noto anche come sintesi, si basa su tecniche di *deep learning*[117] e alberi decisionali[118]. Le fonti per la generazione dei dati sintetici possono essere: (i)dati reali, (ii)conoscenze raccolte dagli analisti oppure (iii) una combinazione di queste due.

L'utilizzo dei dati sintetici può migliorare la privacy nelle tecnologie adottando un approccio basato sulla protezione dei dati fin dalla progettazione, e potrebbe contribuire ad attenuare i *bias* nel trattamento creando modelli equi di intelligenza artificiale caratterizzati da una maggiore rappresentatività delle minoranze.[119]

I dati sintetici, dunque, si sviluppano su due direzioni: l'emulazione di informazioni chiave presenti in *dataset* reali garantendo la privacy e la

[117] Sottoinsieme del machine learning che utilizza reti neurali multilivello per simulare il complesso potere decisionale del cervello umano

[118] Un grafo di decisioni e delle loro possibili conseguenze

[119] https://www.edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en?etrans=it

creazione di scenari di test differenti per valutare fenomeni non coperti dai dati disponibili.[120]

***BayesBoost***

BayesBoost è un metodo per identificare e correggere i bias nei dati, utilizzando avanzate tecniche di generazione di dati sintetici basate su reti bayesane e tecniche di boosting. È stato progettato specificatamente per affrontare problemi di sottorappresentazione di determinati gruppi nei dataset che possono influire sull'equità ed efficacia dei modelli predittivi.[121]

Nello studio in questione, viene creato un dataset appositamente non equo, poi si seleziona un attributo protetto (in questo caso etnia) ed una variabile target, in questo caso la malattia da predire è l'infarto.

Successivamente, attraverso un'analisi dell'incertezza, si utilizza un modello di classificazione per individuare campioni difficili da prevedere. Questi campioni, quindi, raffigureranno i gruppi sottorappresentati.

A questo punto, vengono generati dati sintetici basati su dati originali che terranno conto delle caratteristiche dei gruppi sottorappresentati e genereranno nuovi campioni che equilibrano il dataset, riflettendo una distribuzione proporzionale dei gruppi più equa.

Lo studio, indipendente dalla azienda creatrice del prodotto[122], ha dimostrato che, innanzitutto, Bayesboost è stato in grado di identificare i

[120] J. Baumann, A. Castelnovo, R. Crupi, N. Inverardi e D. Regoli "Bias on Demand: A Modelling Framework That Generates Synthetic Data with Bias" 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT'23) Chicago, IL

[121] B.Draghi, Z.Wang, P.Myles, A. Tucker "Identifying and handling data bias within primary healthcare data using synthetic data generators" Helyon, vol.10, 2024, e24164

[122] Ivi p 10

gruppi sottorappresentati e segnalare sia la necessità di integrare dati su determinati gruppi che la non necessità di farlo per altri.
Inoltre, il sistema è stato in grado di generare correttamente i dati necessari per equilibrare il dataset, riducendo i bias ed ha effettivamente incrementato l'accuratezza del modello predittivo.
Tuttavia, BayesBoost presenta diversi limiti. Innanzitutto, è in grado di gestire un solo attributo protetto per volta creando problemi nel caso in cui più attributi debbano essere presi in considerazione, inoltre è più efficace con attributi categorici[123] piuttosto che con attributi continui[124].
Infine, al momento il modello è ottimizzato solo per le classificazioni binarie, come ad esempio: sì/no, malato/non malato.[125]

[123] Variabili che possono assumere un numero limitato di valori distinti. Ad esempio: Genere (maschio e femmina) ed Etnia(Caucasico, Africano, Asiatico)
[124] Variabili che possono assumere qualsiasi valore numerico all'interno di un intervallo. Ad esempio: Età (0-100) anni o peso (0- 100 kg)
[125] B.Draghi, Z.Wang, P.Myles, A. Tucker "Identifying and handling data bias within primary healthcare data using synthetic data generators" *Cit. p 10*

## 3.3 European Health Data Space

L'European Health Data Space (EHDS), Spazio Europeo dei Dati Sanitari, è una proposta di regolamento approvata dal parlamento europeo nell'aprile del 2024.[126]

L'EHDS è il prodotto di una strategia europea per i dati che mira a creare un mercato unico dei dati, assicurando la competitività globale e la sovranità digitale dell'Unione Europea. Si tratta di un progetto ambizioso che punta a mantenere una garanzia sul controllo dei propri dati da parte dei cittadini e delle aziende, permettendo allo stesso tempo una maggiore disponibilità di dati per l'uso nell'economia e nella società.[127]

L'EHDS intende infatti creare uno spazio europeo dei dati sanitari, una vera e propria piattaforma centrale per la sanità digitale chiamata "MyHealth@EU".[128]

Di particolare rilevanza ai fini dei bias nel trattamento dei dati sanitari, sarà l'obiettivo del regolamento di porre nuove basi giuridiche per l'uso secondario dei dati sanitari. Per uso "secondario dei dati" si intende il trattamento dei dati sanitari elettronici per le finalità di cui al capo IV, articolo 34 del regolamento. Tra i dati utilizzati possono figurare sia i dati sanitari elettronici personali originariamente raccolti nel contesto dell'uso primario, ma anche dati sanitari elettronici raccolti ai fini dell'uso secondario.[129]

---

[126] Proposta di REGOLAMENTO DEL PARLAMENTO EUROPEO E DEL CONSIGLIO sullo spazio europeo dei dati sanitari

[127] https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en

[128] *Proposta di regolamento EHDS* Art. 12

[129] Proposta di regolamento EHDS Art.2(e)

Il capo IV del regolamento, infatti, riguarderà le condizioni generali relative all'uso secondario dei dati sanitari elettronici. L'articolo 33 elenca le categorie minime di dati per l'uso secondario mentre l'articolo 34 definisce le finalità per le quali è possibile trattare i dati includendo, per esempio, motivi di pubblico interesse nell'ambito della sanità pubblica o medicina del lavoro[130] o l'attività di ricerca scientifica nel settore sanitario o dell'assistenza.[131]

Per noi di particolare rilevanza è il punto (g) che consentirebbe l'uso secondario dei dati per "attività di addestramento, prova e valutazione degli algoritmi anche nell'ambito di dispositivi medici, sistemi di IA e applicazioni di sanità digitale, che contribuiscono alla sanità pubblica o alla sicurezza sociale, oppure che garantiscono elevati livelli di qualità e sicurezza dell'assistenza sanitaria, dei medicinali o dei dispositivi medici"[132]

Questo punto permetterebbe di addestrare l'intelligenza artificiale con una quantità e qualità maggiore di dati, creando dataset qualitativi, che come dimostrato sia nel capitolo 1.2 che nel capitolo 3.1, comporterebbe una diminuzione significativa dei bias nell'analisi e produzione dei risultati.

Inoltre, il punto "h" dell'articolo 34, pone un ulteriore finalità possibile per l'uso secondario dei dati. In particolare, permetterebbe l'utilizzo per l'erogazione di un'assistenza sanitaria *personalizzata* che consiste nel valutare, mantenere o ripristinare lo stato di salute delle persone fisiche sulla base dei dati sanitari o di altre persone fisiche. Questo articolo è

[130] Proposta di regolamento EHDS Art. 34(a)
[131] Proposta di regolamento EHDS Art.34(e)
[132] Proposta di regolamento EHDS Art. 34(g)

importante per quanto riguarda la cd. “Precision Health”, ovvero il concetto di assistenza sanitaria personalizzata basata su dati genetici, genomici e omici di ogni individuo, integrati con fattori sociali, economici, culturali e ambientali.[133]

Infatti, in risposta ai metodi tradizionali di diagnosi e trattamento basati su sintomi clinici, che possono variare tra etnie portando ad errori o disparità, lo studio MAGIC illustra come l’uso di “biomarcatori”[134] per la malattia *graft-versus-host* (GVHD)[135], permette una valutazione più oggettiva e universale dei pazienti, riducendo l’impatto delle differenze etniche o altre caratteristiche demografiche.[136]

Tornando alla proposta di regolamento, una forma di tutela contro le possibili discriminazioni o abusi dall’uso secondario dei dati sanitari, è enunciata nell’articolo 35. In particolare il punto “a” vieta l’uso secondario dei dati se volti ad “adottare decisioni pregiudizievoli per una persona fisica sulla base dei suoi dati sanitari elettronici; per essere considerate "decisioni", esse devono produrre effetti giuridici o incidere in modo analogo significativamente su tali persone fisiche”[137]; mentre il punto “b” ne vieterebbe l’uso per “adottare decisioni, in relazione a una persona fisica o a gruppi di persone fisiche, al fine di escluderle dal beneficio di un

[133] Panesar, Arjun. "What Is Precision Health?" In Precision Health and Artificial Intelligence: With Privacy, Ethics, Bias, Health Equity, Best Practices, and Case Studies, 19-45. Apress, 2023

[134] Un biomarcatore è un indicatore biologico, come una sequenza di DNA o una proteina, che è correlato con una data malattia o con una risposta a un determinato trattamento.

[135] Reazione immunitaria esercitata dalle cellule trapiantate (provenienti dal donatore) nei confronti dei tessuti della persona che le riceve(ricevente). La frequenza e gravità aumentano in base al grado di diversità genetica tra donatore e ricevente.

[136] Hamad, Nada. "Precision Medicine May Mitigate Racial Biases." Blood, vol. 144, no. 9, 2024, pp. 927-929

[137] Proposta di regolamento EHDS Art 35(a)

contratto di assicurazione o di modificare i loro contributi e premi assicurativi".[138]

Il regolamento sembrerebbe fare molti passi avanti per quanto riguarda la promozione di una ricerca accurata, qualitativa e rappresentativa. Tuttavia, sembrerebbe che la proposta possa comportare rischi significativi per le libertà e diritti delle persone.[139]

Innanzitutto, sembrano abbassarsi i requisiti di trasparenza e consenso rispetto al GDPR, riducendo la capacità dei pazienti di sapere chi utilizza i loro dati e per quali scopi, considerato che l'uso secondario dei dati è consentito quasi per default, con una supervisione limitata da parte delle autorità pubbliche. Un secondo rischio riguarderebbe la creazione di dataset così grandi, che, per quanto possano beneficiare sulla ricerca, aumenterebbero la possibilità di re-identificare le persone, compromettendo la loro privacy.[140]

L'uso estensivo dei dati senza un consenso adeguato potrebbe portare a una perdita di fiducia nei confronti della ricerca e delle istituzioni, per questo motivo sarebbe ideale che i rischi vengano mitigati allineando il regolamento con le normative europee già esistenti (come il GDPR), limitando l'uso dei dati senza consenso esplicito e rafforzando la supervisione da parte degli organismi nazionali.[141]

---

[138] Proposta di regolamento EHDS Art 35(b)

[139] Marelli L, Stevens M, Sharon T, Van Hoyweghen I, Boeckhout M, Colussi I, Degelsegger-Márquez A, El-Sayed S, Hoeyer K, van Kessel R, Zając DK, Matei M, Roda S, Prainsack B, Schlünder I, Shabani M, Southerington T. The European health data space: Too big to succeed? Health Policy. 2023 p.1-4 Sep;135:104861. doi: 10.1016/j.healthpol.2023.104861. Epub 2023 Jun 26. PMID: 37399677; PMCID: PMC10448378.

[140] Ivi p 2

[141] Ivi p 1

# *CONCLUSIONE*

L'analisi condotta in questa tesi ha evidenziato la complessità del fenomeno dei bias nel trattamento dei dati sanitari, un problema di natura tecnica ma profondamente intrecciato con questioni sia etiche che giuridiche.

È stato rilevato come il bias possa emergere in ogni fase del ciclo di vita del dato, dalla raccolta, all'analisi, all'interpretazione del risultato. Sono state poi individuate le implicazioni negative sulla ricerca medica e sull'accesso a servizi sanitari equi e qualitativi.

Di seguito, si è analizzato come il quadro normativo attuale, con particolare riferimento al GDPR e l'AI Act, sebbene offra sicuramente importanti strumenti di protezione, come la classificazione dei dati sanitari come categorie particolari e i requisiti di trasparenza e qualità per i sistemi di intelligenza artificiale, sembra essere ancora lacunoso per una tutela più completa.

Infine, la tesi ha voluto esplorare strategie sia tecniche che giuridiche per la mitigazione dei *bias*, con un focus su soluzioni innovative come i dati sintetici e il ruolo emergente dello Spazio Europeo dei Dati Sanitari (EHDS). Strategie che, per quanto promettenti, è fondamentale vengano accompagnate da un'attenta supervisione.

In conclusione, affrontare il problema dei *bias* nei dati sanitari richiede uno sforzo coordinato e multidisciplinare che combini progresso tecnologico, rigore normativo e sensibilità etica.

Solo in questo modo, una risorsa fondamentale come i dati sanitari potrà contribuire realmente alla creazione di una sanità equa, inclusiva e rispettosa dei diritti fondamentali.

# *BIBLIOGRAFIA*


- Ashrafuzzaman, M., Milu, M., Anjum, A., Khanam, F., & Md. A. R. (2022). Chapter 5 - *Big data analytics techniques for healthcare*. In: Keikhosrokiani, B. P. (ed.) *Big Data Analytics for Healthcare. Datasets, Techniques, Life Cycles, Management, and Applications* (pp. 49-62). New York: Academic Press.
- Bartholomeeusen S, Kim CY, Mertens R, Faes C, Buntinx F. *The denominator in general practice, a new approach from the Intego database.* Fam Pract. 2005 Aug;22(4):442-7. doi: 10.1093/fampra/cmi054. Epub 2005 Jun 17. PMID: 15964863.
- Draghi B, Wang Z, Myles P, Tucker A. *Identifying and handling data bias within primary healthcare data using synthetic data generators.* Heliyon. 2024 Jan 10;10(2): e24164. doi: 10.1016/j.heliyon. 2024.e24164. PMID: 38288010; PMCID: PMC10823075.
- Busuioc M, Curtin D, Almada M. *Reclaiming transparency: contesting the logics of secrecy within the AI Act. European Law Open.* 2023;2(1):79-105. doi:10.1017/elo.2022.47
- Carrera PM, Lambooij MS. *Implementation of Out-of-Office Blood Pressure Monitoring in the Netherlands: From Clinical Guidelines to Patients' Adoption of Innovation*. Medicine (Baltimore). 2015 Oct;94(43):e1813. doi: 10.1097/MD.0000000000001813. PMID: 26512579; PMCID: PMC4985393.

- Daneshjou, R., Smith, M. P., Sun, M. D., Rotemberg, V., & Zou, J. (2021). *Lack of Transparency and Potential Bias in Artificial Intelligence Data Sets and Algorithms: A Scoping Review.* JAMA Dermatology, 157(11), 1362–1369
- Feng Chen, Liqin Wang, Julie Hong, Jiaqi Jiang, Li Zhou. *Unmasking bias in artificial intelligence: a systematic review of bias detection and mitigation strategies in electronic health record-based models*. Journal of the American Medical Informatics Association. Volume 31, Issue 5, May 2024, Pages 1172–1183.
- G. Manikandan, S. Abirami, K. Gokul, G. Deepalakshmi. *Chapter 1: Big data analytics in healthcare theory, tools, techniques and its applications.* Elsevier, 2022.
- Hamad, Nada. *"Precision Medicine May Mitigate Racial Biases."* Blood, vol. 144, no. 9, 2024, pp. 927-929
- J. Baumann, A. Castelnovo, R. Crupi, N. Inverardi, and D. Regoli. 2023. *Bias on Demand: A Modelling Framework That Generates Synthetic Data With Bias.* In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23). Association for Computing Machinery, New York, NY, USA, 1002–1013. https://doi.org/10.1145/3593013.3594058
- Lau, P.L. (2024). *AI Gender Biases in Women's Healthcare: Perspectives from the United Kingdom and the European Legal Space*. In: Gill-Pedro, E., Moberg, A. (eds) YSEC Yearbook of

Socio-Economic Constitutions 2023. YSEC Yearbook of Socio-Economic Constitutions, vol 2023. Springer, Cham.

- Malgieri, G., & Comandé, G. (2017*). Sensitive-by-distance: quasi-health data in the algorithmic era. Information & Communications Technology* Law, 26(3), 229–249
- Marelli, Luca, Marthe Stevens, Tamar Sharon, Ine Van Hoyweghen, Martin Boeckhout, Ilaria Colussi, et al. *"The European Health Data Space: Too Big to Succeed?"* Health Policy 135 (2023): 104861
- Marvin van Bekkum, Frederik Zuiderveen Borgesius. *Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception. "Computer Law and Security Review"* V.48. p.105-770. 2023
- Melissa Hall, Laurens van der Maaten, Laura Gustafson, Maxwell Jones, Aaron Adcock. *A Systematic Study of Bias Amplification.* Cornell University, 2022
- Moura L, Furuie SS, Gutierrez MA, Tachinardi U, Rebelo MS, Alcocer P, Melo CP. *Lossy compression techniques, medical images, and the clinician*. MD Comput. 1996 Mar-Apr;13(2):155-9, 172. PMID: 8684278.
- *Kaidbey KH, Agin PP, Sayre RM, Kligman AM. Photoprotection by melanin—a comparison of black and Caucasian skin. J Am Acad Dermatol. 1979 Sep;1(3):249-60. doi: 10.1016/s0190-9622(79)70018-1. PMID: 512075.*

- Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. *Addressing bias in big data and AI for health care: A call for open science.* Patterns (N Y). 2021 Oct 8;2(10):100347. doi: 10.1016/j.patter.2021.100347. PMID: 34693373; PMCID: PMC8515002.
- *Oxford English Dictionary, Oxford, OUP, 2023.*
- Panesar, Arjun. *"What Is Precision Health?" In Precision Health and Artificial Intelligence: With Privacy, Ethics, Bias, Health Equity, Best Practices, and Case Studies,* 19-45. Apress, 2023
- Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, Boccia S. *Benefits and challenges of Big Data in healthcare: an overview of the European initiatives.* Eur J Public Health. 2019 Oct 1;29(Supplement_3):23-27. doi: 10.1093/eurpub/ckz168. PMID: 31738444; PMCID: PMC6859509.
- Popovic A, Huecker MR. *Study Bias.* 2023 Jun 20. In: StatPears [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan–. PMID: 34662027.
- Quinn P, Malgieri G. *The Difficulty of Defining Sensitive Data—The Concept of Sensitive Data in the EU Data Protection Framework.* German Law Journal. 2021;22(8):1583-1612. doi:10.1017/glj.2021.79
- Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, Mohr DC. *Mobile Phone Sensor Correlates of Depressive Symptom*

*Severity in Daily-Life Behavior: An Exploratory Study*. J Med Internet Res. 2015 Jul 15

- Vayena, E., Gasser, U. (2016). *"Strictly Biomedical? Sketching the Ethics of the Big Data Ecosystem in Biomedicine".* In: Mittelstadt, B., Floridi, L. (eds) The Ethics of Biomedical Big Data. Law, Governance and Technology Series, vol 29. Springer, Cham.
- Van der Bij S, Khan N, Ten Veen P, de Bakker DH, Verheij RA. *Improving the quality of EHR recording in primary care: a data quality feedback tool*. J Am Med Inform Assoc. 2017 Jan;24(1):81-87
- *Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. J Med Internet Res. 2018 May 29;20(5):e185. doi: 10.2196/jmir.9134. PMID: 29844010; PMCID: PMC5997930.*
- *Wilkinson C, Bebb O, Dondo TB, et al. Sex differences in quality indicator attainment for myocardial infarction: a nationwide cohort study. Heart 2019; 105:516-523*

# *ATTI NORMATIVI*

- *Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale)*
- *Proposta di REGOLAMENTO DEL PARLAMENTO EUROPEO E DEL CONSIGLIO sullo spazio europeo dei dati sanitari*
- *Regolamento (UE) 2018/1725 del Parlamento europeo e del Consiglio, del 23 ottobre 2018, sulla tutela delle persone fisiche in relazione al trattamento dei dati personali da parte delle istituzioni, degli organi e degli organismi dell'Unione e sulla libera circolazione di tali dati, e che abroga il regolamento (CE) n. 45/2001 e la decisione n. 1247/2002/CE*
- *Regolamento di esecuzione (UE) 2018/679 della Commissione, del 3 maggio 2018, che rinnova l'approvazione della sostanza attiva forchlorfenuron in conformità al regolamento (CE) n. 1107/2009 del Parlamento europeo e del Consiglio relativo all'immissione sul mercato dei prodotti fitosanitari e che modifica l'allegato del regolamento di esecuzione (UE) n. 540/2011 della Commissione*
- *Regolamento (UE) 2017/746 del Parlamento europeo e del Consiglio, del 5 aprile 2017, relativo ai dispositivi medico-diagnostici in vitro e che abroga la direttiva 98/79/CE e la decisione 2010/227/UE della Commissione (Testo rilevante ai fini del SEE. )*
- *Regolamento (UE) 2017/745 del Parlamento europeo e del Consiglio, del 5 aprile 2017, relativo ai dispositivi medici, che modifica la direttiva 2001/83/CE, il regolamento (CE) n. 178/2002 e il regolamento (CE) n. 1223/2009 e che abroga le direttive 90/385/CEE e 93/42/CEE del Consiglio (Testo rilevante ai fini del SEE.*

- *DIRETTIVA (UE) 2016/680 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 27 aprile 2016 relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati e che abroga la decisione quadro 2008/977/GAI del Consiglio*
- *REGOLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati)*
- *Carta dei diritti fondamentali dell'Unione Europea*
- *Direttiva 95/46/CE del Parlamento europeo e del Consiglio, del 24 ottobre 1995, relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati*
- *Convenzione sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale Strasburgo, 28 gennaio 1981*

# *SITOGRAFIA*


- Agenda digitale (12 marzo 2024) IA emotiva e bias algoritmici: l'impatto nel settore sanitario.
  Disponibile su:
  https://www.agendadigitale.eu/sanita/ia-emotiva-e-bias-algoritmici-limpatto-nel-settore-sanitario

- Commissione Europea (1° agosto 2024) Entra in vigore il regolamento sull'IA.
  Disponibile su:
  https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_it

- Coursera (20 novembre 2024) What is Data wrangling? Definition, Steps and why it matters.
  Disponibile su:
  https://www.coursera.org/articles/data-wrangling

- Edps Europa eu (nd) Synthetic Data.
  Disponibile su:
  https://www.edps.europa.eu/presspublications/publications/techsonar/syntheticdata_en#:~:text=Synthetic%20data%20is%20artificial%20data%20that%20is%20generated,similar%20results%20when%20undergoing%20the%20same%20statistical%20analysis.

- European parlament (23 marzo 2023) Big data: definizione, benefici e sfide.
  Disponibile su: https://www.europarl.europa.eu/topics/it/article/20210211STO97614/big-data-definizione-benefici-e-sfide

- Pharmaceutical Journal (23 agosto 2021) Pharmacies to get up to £1,800 in extra contract funding for providing heart checks.
  Disponibile su: https://pharmaceutical-journal.com/article/news/pharmacies-to-get-up-to-1800-in-extra-contract-funding-for-providing-heart-checks

- Policy Lab (22 marzo 2024) Fair decision-making: Can humans save us from biased AI?
  Disponibile su: https://policy-lab.ec.europa.eu/news/fair-decision-making-can-humans-save-us-biased-ai-2024-03-22_en

- Protezione dati personali (11 ottobre 2023) Regolamento generale per la protezione dei dati.
  Disponibile su: https://protezionedatipersonali.it/regolamento-generale-protezione-dati