

Ripensare l'evidenza: approccio multidisciplinare all'uso dei dati sintetici in ambito sanitario – Parte I

*Alessandro Brusadelli, Davide Ruffo, Chiara di Somma, Alexander Boudewijn**

RETHINKING EVIDENCE: A MULTIDISCIPLINARY APPROACH TO THE USE OF SYNTHETIC DATA IN HEALTHCARE - PART I

ABSTRACT: This essay explores the role of synthetic data in clinical trials and Health Technology Assessment, highlighting its transformative potential on epistemological and legal levels. Adopting a multidisciplinary approach, it examines the compatibility of such data with European health law and the evidence-based design paradigm. The aim is to propose a systemic interpretation of the current legal framework that legitimizes synthesization as an emerging form of evidence generation, calling for the development of a new legal grammar capable of reconciling technological innovation, fundamental rights, and the sustainability of healthcare systems.

KEYWORDS: Synthetic data; clinical trials; Health Technology Assessment (HTA); artificial intelligence; data protection

ABSTRACT: Il saggio esplora il ruolo dei dati sintetici nella sperimentazione clinica e nell'Health Technology Assessment, evidenziandone il potenziale trasformativo sul piano epistemologico e giuridico. Muovendo da un approccio multidisciplinare, si analizza la compatibilità di tali dati con il diritto europeo della salute e il paradigma dell'evidence-based design. L'obiettivo è proporre una lettura sistemica dell'ordinamento vigente che legittimi la sintetizzazione come forma emergente di costruzione dell'evidenza, sollecitando l'elaborazione di una nuova grammatica giuridica

**Alessandro Brusadelli: Dottorando in Intelligenza Artificiale in Medicina e Innovazione nella Ricerca Clinica e Metodologica presso il Dipartimento di Scienze Cliniche e Sperimentali dell'Università degli Studi di Brescia. L'autore ha curato: Introduzione; Oltre il soggetto: identità computazione e crisi della referenzialità nel diritto dei dati personali; Evidence by design: la funzione orientante del Regolamento HTA nel governo della sperimentazione; Un primo bilancio. L'autore inoltre è co-responsabile della curatela editoriale dell'intero saggio (Parte I e Parte II). Mail: alessandro.brusadelli@unibs.it. Davide Ruffo: Avvocato, Cultore della Materia in Diritto Comunitario presso l'Università Europea di Roma, Chief Institutional Officer AINDO. L'autore ha curato: Il Disegno di Legge n. 1146/2024 sull'intelligenza artificiale: analisi critica dell'articolo 8 e considerazioni specifiche sull'articolo 9; Un primo bilancio. Mail: davide.ruffo@aindo.com. Chiara di Somma: Avvocato, dottore di ricerca in diritto pubblico e costituzionale, funzionario del Garante per la protezione dei dati personali. L'autore ha curato: L'evoluzione delle condizioni di liceità per il trattamento di dati personali per scopi di ricerca medica. Il contributo dell'autore è redatto a titolo personale, quanto ivi riportato e le opinioni ivi espresse impegnano solo l'autore. Mail: cdisomma@luiss.it. Alexander Boudewijn: PhD, Data Privacy Researcher, AINDO. L'autore ha curato: I dati sintetici come paradigma emergente per la valutazione delle tecnologie sanitarie (HTA): dalla protezione alla valorizzazione del dato; Il contributo dei dati sintetici alla disponibilità e interoperabilità dell'evidenza; Simulazione e confronto: il ruolo dei dati sintetici nella valutazione comparativa; Verso una nuova grammatica dell'HTA basata sull'evidenza. Mail: alexander@aindo.com. Contributo sottoposto a doppio referaggio anonimo.*

capace di coniugare innovazione tecnologica, diritti fondamentali e sostenibilità dei sistemi sanitari.

PAROLE CHIAVE: Dati sintetici; sperimentazione clinica; valutazione delle tecnologie sanitarie (HTA); intelligenza artificiale; protezione dei dati personali

SOMMARIO: 1. Introduzione - 2. La disciplina giuridica dei dati sintetici: tra protezione dei dati personali e promozione della ricerca - 2.1. Oltre il soggetto: identità computazionale e crisi della referenzialità nel diritto dei dati personali - 2.2. Il Disegno di Legge n. 1146/2024 sull'intelligenza artificiale: analisi critica dell'articolo 8 e considerazioni specifiche sull'articolo 9 - 2.3. L'evoluzione delle condizioni di liceità per il trattamento di dati personali per scopi di ricerca medica - 2.4. *Evidence by design*: la funzione orientante del Regolamento HTA nel governo della sperimentazione - 2.5. Un primo bilancio - 3. I dati sintetici come paradigma emergente per la valutazione delle tecnologie sanitarie (HTA): dalla protezione alla valorizzazione del dato - 3.1. Il contributo dei dati sintetici alla disponibilità e interoperabilità dell'evidenza - 3.2. Simulazione e confronto: il ruolo dei dati sintetici nella valutazione comparativa - 3.3. Verso una nuova grammatica dell'HTA basata sull'evidenza.

1. Introduzione

Nel contesto contemporaneo della ricerca biomedica, il paradigma sperimentale tradizionale si confronta con una crescente tensione sistemica tra, da un lato, l'esigenza di garantire un accesso tempestivo e sicuro a nuove terapie, e, dall'altro, il rispetto rigoroso dei principi etici, di sicurezza e protezione dei dati personali. Tale tensione non si traduce in una semplice dialettica tra progresso e cautela, ma si colloca in una più profonda riconfigurazione epistemologica del rapporto tra dato, evidenza e regolazione. È in questa cornice che si staglia l'emergere dei dati sintetici, non come mera innovazione tecnica, ma come tecnologia capace di ridisegnare le premesse metodologiche e normative della sperimentazione clinica. La sintetizzazione introduce una novità significativa: essa consente la generazione, attraverso processi algoritmici di modellizzazione statistica, di nuovi insiemi di dati disancorati da referenti empirici specificamente individuati e potenzialmente idonei a rappresentare popolazioni cliniche verosimili. In tal modo, il dato sintetico non si configura più come derivazione del paziente fisico, bensì come espressione computazionale di pazienti possibili, legittimando nuove forme di costruzione dell'evidenza scientifica in grado di superare i vincoli strutturali della sperimentazione convenzionale, quali la scarsità di dati (ad esempio nel caso di malattie rare), la durata dei trial, la difficoltà di reclutamento dei pazienti, senza compromettere la qualità dell'evidenza scientifica originaria.

Alla luce di quanto sopra, il presente pionieristico lavoro si propone di interrogare il significato giuridico, etico e regolatorio dell'impiego dei dati sintetici nella sperimentazione clinica, muovendo dall'ipotesi che tale innovazione tecnologica possa assumere una funzione sistemica all'interno del diritto europeo della salute. In particolare, si analizzerà la potenziale compatibilità dei dati sintetici con il Regolamento (UE) 2021/2282 sull'Health Technology Assessment (HTA), quale espressione di una razionalità regolativa fondata sull'evidence-based design, e con l'evoluzione dei presupposti di liceità del trattamento dei dati personali per finalità di ricerca scientifica. L'obiettivo non è quello di rivendicare per i dati sintetici una disciplina autonoma, quanto piuttosto di esplorarne la possibile inclusione in un'interpretazione evolutiva e sistemica dell'ordinamento vigente, orientata dai principi di efficacia, sostenibilità, equità e innovazione.



In questa prospettiva, la legittimazione giuridica dei dati sintetici nella pratica valutativa non rappresenta una deroga, ma un'espressione coerente della trasformazione paradigmatica che investe il nesso tra tecnologia, diritto e conoscenza scientifica.

Il lavoro si articola in due contributi distinti ma consecutivi e concettualmente e tematicamente collegati, all'interno dei quali si affronta l'argomento in oggetto da prospettive differenti in costante dialogo tra loro: da quella giuridica a quella tecnica, da quella clinica a quella economica, non tralasciando le decisive riflessioni elaborate dall'etica e dalla filosofia teoretica. È precisamente in questa dialettica inter-transdisciplinare, resa tale anche dalla variegata eterogeneità delle competenze specifiche dei diversi contributori e l'unica in grado di restituire l'interesse e la complessità del fenomeno dei dati sintetici, che risiede l'aspetto pionieristico del presente lavoro.

Il primo contributo è dedicato al complesso quadro normativo all'interno del quale si colloca tale tecnologia emergente. Nello specifico, il primo capitolo affronta il nodo concettuale e giuridico della qualificazione del dato sintetico, evidenziando la tensione tra protezione della persona e la promozione della ricerca scientifica nell'evoluzione del diritto dei dati. Il secondo capitolo, alla luce dell'innovazione tecnologica dei dati sintetici quale paradigma emergente per la valutazione delle tecnologie sanitarie, propone una rilettura del Regolamento HTA come dispositivo di orientamento normativo nella costruzione dell'evidenza clinica.

Il secondo contributo è incentrato sulle implicazioni cliniche, economiche, etiche e filosofiche dei dati sintetici. In particolare, il primo capitolo, anche attraverso l'analisi di un caso di studio, prende in considerazione le trasformazioni strutturali nei modelli di sperimentazione indotte dall'impiego di IA e dati sintetici, dedicando una riflessione alle implicazioni etiche. Il secondo capitolo considera le ricadute economiche e organizzative dell'impiego dei dati sintetici, con riferimento ai modelli di valutazione, rimborso e sostenibilità nei sistemi sanitari. Nella sezione conclusiva, facendosi supportare dalle elaborazioni concettuali della filosofia teoretica, si vuole proporre l'ambiziosa tesi secondo cui non solo i dati sintetici possono essere validi da un punto di vista scientifico così come i dati clinici di partenza, ma addirittura si muovono sullo stesso piano di realtà di quest'ultimi, andando così a scardinare la convinzione del senso comune che sia reale solo il dato clinico e non quello sintetico.

Attraverso questo percorso, il saggio inteso nella sua complessità intende offrire una ricostruzione critica del ruolo dei dati sintetici nel governo dell'innovazione biomedica, prospettando al contempo l'urgenza di una nuova e inedita grammatica giuridica capace di accoglierne la portata trasformativa senza sacrificare l'imprescindibile funzione garantista del diritto. In tal senso, il dato sintetico non è solo un artefatto tecnico-tecnologico, ma una nuova entità che impone al diritto, e non solo, di ripensare le proprie categorie.

*Is
Law*

2. La disciplina giuridica dei dati sintetici: tra protezione dei dati personali e promozione della ricerca

2.1. Oltre il soggetto: identità computazionale e crisi della referenzialità nel diritto dei dati personali

L'irruzione della digitalizzazione nel tessuto della società ha profondamente inciso sullo statuto ontologico e giuridico dell'identità personale, costringendo il diritto a confrontarsi con forme inedite di soggettività¹. In tale contesto l'identità non si esaurisce più nella corporeità tangibile né nella dimensione relazionale immediata, ma si moltiplica in una costellazione di proiezioni informazionali, distribuite nei circuiti della comunicazione digitale, nei database, nei modelli predittivi e negli ambienti algoritmici². Il soggetto, così come emerge nell'era dell'intelligenza artificiale (d'ora in poi IA), è al tempo stesso origine e prodotto di processi computazionali che ne tracciano profili, prefigurano comportamenti, anticipano decisioni³. Tale evoluzione impone una rilettura delle categorie giuridiche classiche, a partire da quella d'identità, che, nel suo fondamento costituzionale, si configura quale principio immanente al riconoscimento della persona come centro di imputazione di diritti e doveri (art. 2 Cost.) e, più in particolare, come diritto a preservare la coerenza, la continuità e l'integrità della propria rappresentazione sociale⁴. In questa prospettiva, la protezione dei dati personali assurge a strumento di garanzia della persona nella sua dimensione più profonda: non soltanto come mezzo di autodeterminazione informativa, ma come presidio dell'identità in una società dove l'io giuridico si costruisce, si frammenta e si ricompone nei linguaggi dell'informazione digitale⁵.

La progressiva smaterializzazione dell'identità del soggetto ha trovato un riscontro normativo anticipatorio e strutturato nella disciplina della protezione dei dati personali. Quest'ultima, originariamente ancorata a un paradigma difensivo della riservatezza della persona intesa quale ambito privato inviolabile, si è progressivamente riconfigurata in una architettura normativa più complessa, fondata sul principio di responsabilizzazione (accountability)⁶ e sul governo trasparente dei flussi informativi. Il Regolamento UE 2016/679, comunemente noto con l'acronimo inglese GDPR (General Data Protection Regulation),

¹ L. FLORIDI, *The Onlife Manifesto. Being Human in a Hyperconnected Era*, Cham, 2014. DOI: 10.1007/978-3-319-04093-6.

² S. ZUBOFF, *capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Milano, 2020.

³ A. ROUVROY, *The End(s) of Critique: Data Behaviourism versus Due Process*, in M. HILDEBRANDT, K. DE VRIES (a cura di), *Privacy, Due Process and the Computational Turn*, Londra, 2013.

⁴ Per il legame tra identità e coerenza narrativa si veda P. RICOEUR, *Soi-même comme un autre*, Paris, 1990.

⁵ Per il fondamento relazionale e riconoscitivo dell'identità si veda J. BUTLER, *Giving an Account of Oneself*, New York, 2005.

⁶ Il principio di accountability, sancito all'art. 5, par. 2 del GDPR, segna una discontinuità epistemica nel diritto della protezione dei dati, ridefinendo il ruolo del titolare del trattamento non più come semplice destinatario di obblighi normativi, ma come soggetto attivamente responsabile della conformità del trattamento ai principi fondamentali del Regolamento. Tale principio, riconosciuto dalla dottrina quale perno dell'intera architettura regolatoria europea, introduce un'inversione strutturale dell'onere della prova, fondando un regime di responsabilizzazione sostanziale che attraversa l'intero ciclo vitale del dato. Esso esige l'adozione di misure tecniche, organizzative e documentali proporzionate al rischio e ispirate alla logica della prevenzione, secondo il paradigma della data protection by design and by default, configurando così una forma giuridica della vigilanza informata e della governance del dato. A tal proposito si rimanda a G. Scorza, *Il principio di accountability: uno dei pilastri del GDPR*, in "Altalex", 13 febbraio 2018, disponibile su: <https://www.altalex.com/documents/news/2018/02/13/il-principio-di-accountability-uno-dei-pilastri-del-gdpr>.



espressione matura di tale evoluzione, riconosce esplicitamente la centralità giuridica dell'identità digitale, tutelandola attraverso la nozione cardine di "dato personale", definito all'art. 4, n. 1 come *"qualsiasi informazione riguardante una persona fisica identificata o identificabile"*⁷. In questa prospettiva, il dato non è un mero supporto informativo, né un elemento ancillare dell'esperienza giuridica del soggetto: esso si configura come parte costitutiva dell'identità personale, una sua proiezione semantica, un frammento che, se manipolato, trasmesso o profilato, ha effetti concreti sulla rappresentazione e sull'autodeterminazione dell'individuo. La disciplina giuridica del trattamento dei dati personali assume, pertanto, un rilievo sistemico: essa non regola semplicemente un insieme di operazioni tecniche, ma diventa presidio dei diritti fondamentali della persona nella società dell'informazione.

L'emergere dei dati sintetici, vale a dire di dati generati da modelli di IA allenati su insiemi di dati raccolti, suggerisce una potenziale revisione del paradigma che sinora ha ancorato la disciplina del trattamento informativo alla referenzialità empirica del dato personale. Il dato sintetico non costituisce più la rappresentazione informazionale di un soggetto fisico determinato, bensì la configurazione statistica di un individuo possibile, frutto di un processo algoritmico di astrazione, modellizzazione e generazione. L'identità, quindi, da presupposto radicato nella relazione tra dato e soggetto, si riconfigura come esito computazionale: non è più l'individuo a generare il dato, ma è l'infrastruttura algoritmica a generare una figura identitaria virtuale, slegata da un referente empirico puntualmente identificato. Tale slittamento comporta una profonda dislocazione semantica e giuridica: si passa dalla rappresentazione digitale della realtà fisica alla produzione simulata di una realtà possibile, dalla tutela della soggettività individuale alla gestione regolatoria di entità modellate da inferenze probabilistiche. Il diritto è dunque chiamato a misurarsi con una mutazione di ordine epistemico, in quanto la discontinuità concettuale introdotta dai dati sintetici sollecita una riflessione critica sulle categorie giuridiche tradizionalmente impiegate per qualificare le diverse forme di dato. L'identità, intesa non più come dato originario da proteggere ma come costruzione sinteticamente prodotta, reclama forse nuove forme di riconoscimento e disciplina?

Se la pseudonimizzazione⁸, ai sensi dell'art. 4, n. 5 GDPR, attenua il rischio di identificabilità mediante la rimozione degli identificatori diretti, e se l'anonimizzazione si propone di rendere strutturalmente irrinunciabile il legame con la persona fisica (sottraendo il dato all'ambito di applicazione del Regolamento), la sintetizzazione segna rottura epistemologica e, come si cercherà di indagare anche normativa. Essa, infatti, non opera per sottrazione di elementi identificativi, ma per generazione autonoma di nuovi dati privi di qualsiasi ancoraggio empirico a soggetti fisici puntualmente individuati⁹. In linea teorica, tali dati potrebbero essere qualificati come informazioni anonime, estranee dunque all'ambito di applicazione del GDPR. Ma la questione resta giuridicamente controversa, specie alla luce della relazione che i dati sintetici possono mantenere con i dataset originari¹⁰. Tale ambivalenza solleva interrogativi sulla soglia di rischio identificativo compatibile con l'esclusione dalla tutela regolatoria, richiedendo un approccio casistico e

⁷ Art. 4, n. 1, GDPR.

⁸ L. BOLOGNINI, C. BISTOLFI, *Pseudonymization and impacts of Big (personal/anonymous) Data processing in the transition from the Directive 95/46/EC to the new EU General Data Protection Regulation*, in *Computer Law & Security Review*, Volume 33, Issue 2, 2017, 171-181. <https://doi.org/10.1016/j.clsr.2016.11.002>

⁹ Garante per la Protezione dei Dati Personali, *Provvedimento del 2 agosto 2024*, n. 477.

¹⁰ E.C. RAFIOTTA, *Intelligenza artificiale e dati sintetici*, in *BioLaw Journal – Rivista di BioDiritto*, n. 3/2024.

prudenziale, anche alla luce della giurisprudenza della Corte di Giustizia dell'Unione Europea in materia di identificabilità potenziale¹¹.

In tale contesto, l'art. 110 del Codice Privacy (d.lgs. 196/2003, come modificato dal d.lgs. 101/2018), che verrà interrogato più analiticamente nei paragrafi successivi, si presenta come uno snodo normativo rilevante, in quanto consente, in presenza di motivate ragioni scientifiche ed etiche, l'utilizzo di dati personali per finalità di ricerca sanitaria anche in assenza del consenso dell'interessato, a condizione che siano adottate misure appropriate indicate dal Garante. In questa prospettiva, i dati sintetici, proprio per la loro struttura non riconducibile a soggetti identificabili, si candidano a rappresentare una risposta normativa e tecnica innovativa: superano le criticità connesse alla disponibilità e trattabilità dei dati raccolti, rispettando al contempo il principio di minimizzazione e garantendo un'elevata qualità scientifica del dato utilizzato. La loro adozione, pertanto, non si limita a un espediente tecnico, ma si configura come potenziale leva di riforma per l'intero sistema di governance dei dati in ambito clinico e sperimentale. Sul piano teorico, l'avvento dei dati sintetici segna una cesura epistemologica di portata sistemica, come si avrà modo di approfondire nelle conclusioni della seconda parte del presente lavoro. La conoscenza, tradizionalmente fondata sull'osservazione diretta della realtà empirica e sulla raccolta esperienziale del dato, si riconfigura come simulazione di mondi possibili, costruiti mediante modelli matematici e processi di generazione algoritmica¹². Il dato non è più impronta dell'esperienza vissuta, ma artefatto computazionale: espressione di un sapere predittivo, potenziale, che affonda le sue radici nella capacità del calcolo di anticipare l'accadere. Una trasformazione che porta con sé ricadute rilevanti sul piano giuridico, in quanto impone un'estensione della previsione normativa oltre i confini dell'identificabilità soggettiva. Non si tratta più di regolare esclusivamente ciò che è personale, ma di interrogarsi su ciò che è computazionalmente generabile, statisticamente plausibile e normativamente rilevante. Il dato sintetico sfida la tenuta delle categorie giuridiche fondate sul binomio soggetto - dato, sollevando interrogativi inediti: quali criteri fondano oggi il riconoscimento giuridico di una rappresentazione? È possibile ed auspicabile tutelare ciò

¹¹ Corte di giustizia dell'Unione europea, *Breyer v Bundesrepublik Deutschland*, causa C-582/14, sentenza del 19 ottobre 2016, ECLI:EU:C:2016:779. Nella sentenza *Breyer v. Bundesrepublik Deutschland* (C-582/14, ECLI:EU:C:2016:779), la Corte di giustizia dell'Unione europea ha articolato una nozione estensiva di dato personale, chiarendo che l'identificabilità dell'interessato non richiede una possibilità diretta e immediata, ma può fondarsi sull'esistenza di mezzi "ragionevolmente utilizzabili" da parte del titolare o di terzi. In tale prospettiva, anche un'informazione apparentemente neutra – come un indirizzo IP dinamico – assume natura personale qualora, in un contesto concreto, risulti tecnicamente e giuridicamente possibile risalire all'identità del soggetto cui si riferisce. La Corte introduce così un criterio fondato sulla concreta accessibilità dell'identificazione, valutata alla luce di elementi quali costi, tempi e disponibilità tecnologiche. La sentenza consolida un'interpretazione estensiva del concetto di dato personale, centrale nell'impianto del GDPR.

¹² A tal proposito si veda P. MIROWSKI, *Machine Dreams. Economics Becomes a Cyborg Science*, Cambridge, 2002. L'autore analizza in chiave storica e teorica il passaggio da una scienza osservativa a una scienza modellizzante e computazionale, mostrando come l'economia, ma più in generale le scienze sociali e naturali, si siano "cyborgizzate", sostituendo l'esperienza empirica con simulazioni algoritmiche. L'opera costituisce una delle più autorevoli riflessioni sul ruolo della computazione come nuovo paradigma epistemico, e risulta centrale per comprendere come i dati sintetici, quali artefatti generativi, incarnino una cesura teorica tra conoscenza esperienziale e sapere predittivo. Si veda anche A. ROUVROY, *op. cit.* Il concetto di "data behaviourism" descrive un paradigma epistemico in cui il diritto perde il proprio radicamento nel soggetto empirico e viene ristrutturato attorno a inferenze algoritmiche sul comportamento. La conoscenza giuridica e sociale non si fonda più sulla descrizione di fatti, ma sulla previsione automatizzata di condotte.

che non deriva da una persona, ma che concorre a costruire rappresentazioni soggettive socialmente ed epistemicamente rilevanti?

Tali interrogativi impongono un ampliamento dello sguardo giuridico, chiamato a misurarsi con la complessità epistemica che l'intersezione tra IA e diritto oggi dischiude. L'irruzione di sistemi computazionali in grado di generare inferenze statistiche e simulazioni probabilistiche non introduce soltanto nuovi strumenti tecnici nel dominio giuridico, ma ne destabilizza i presupposti concettuali, incidendo sul nucleo teorico del diritto stesso.

Il passaggio dalla logica causale e lineare, che ha storicamente strutturato il ragionamento normativo, a una logica di tipo inferenziale, fondata su correlazioni probabilistiche tra elementi eterogenei, segna una discontinuità che travalica il piano delle tecniche regolative. Il giurista è così chiamato a elaborare nuove forme di razionalità giuridica, che siano capaci di dare vita ad uno spazio giuridico in cui soggettività, dati e norme si ricombinino secondo nuove grammatiche ancora in via di definizione.

2.2. Il Disegno di Legge n. 1146/2024 sull'intelligenza artificiale: analisi critica dell'articolo 8 e considerazioni specifiche sull'articolo 9

Se il paragrafo precedente ha messo in evidenza come l'IA solleciti una riconfigurazione profonda delle categorie giuridiche tradizionali, il Disegno di Legge n. 1146/2024 intitolato "Disposizioni e deleghe al Governo in materia di intelligenza artificiale", noto come DDL AI, rappresenta un primo tentativo, ancora iscritto entro coordinate concettuali in evoluzione, di articolare una risposta normativa a tali trasformazioni. Al momento della redazione del presente scritto, il DDL AI si colloca nella fase terminale del proprio iter parlamentare, avendo conseguito l'approvazione del Senato della Repubblica il 20 marzo 2025 e trovandosi ora a transitare nuovamente dinanzi alla medesima Assemblea per una terza lettura, a seguito dell'esame e della votazione presso la Camera dei deputati. La maturità raggiunta dal provvedimento nel suo percorso legislativo, pur essendo imperfetta, unitamente alla lunga gestazione iniziata con la presentazione da parte dell'Esecutivo il 20 maggio 2024, autorizza a ritenere che, almeno con riferimento agli aspetti fondamentali oggetto del presente paragrafo, gli indirizzi e le scelte di fondo possano considerarsi ormai sufficientemente delineati ancorché non consolidati. Questa considerazione trova ulteriore corroborazione nell'analisi dell'architettura normativa delineata dal provvedimento: un'architettura che, lungi dal limitarsi a una mera trasposizione delle prescrizioni europee, tra cui emerge in primo piano l'AI Act¹³, nel diritto interno, manifesta la volontà di configurare un quadro regolatorio autoctono, seppur intenzionalmente coerente e compatibile con il contesto sovranazionale, capace di intercettare le specificità dell'ordinamento nazionale, con particolare attenzione ai settori sensibili della ricerca scientifica e della sanità digitale.

Nel tentativo di costruire un impianto normativo capace di governare le discontinuità introdotte dall'IA, il DDL AI rappresenta dunque un passaggio emblematico verso una ridefinizione delle categorie giuridiche applicabili al trattamento dei dati nell'ambito della ricerca scientifica. L'attenzione riservata dal legislatore ai settori della sanità digitale e dell'innovazione medico-tecnologica segna il riconoscimento della crescente centralità della conoscenza computazionale e della modellizzazione algoritmica nella produzione di evidenza scientifica.

¹³Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale.

È in tale cornice che si colloca l'articolo 8 del DDL AI, il quale introduce, per la prima volta, una disciplina specifica incentrata sull'impiego dell'IA nei processi di ricerca biomedica. Al di là dell'impatto settoriale, la disposizione appare destinata a incidere in profondità sulla grammatica giuridica del trattamento, inaugurando una riflessione sistemica sulla legittimazione giuridica di operazioni computazionali come la sintetizzazione. L'analisi della norma consente quindi di cogliere non solo le tensioni interpretative che ne derivano, ma anche le implicazioni teoriche più ampie per la struttura del diritto della protezione dei dati nell'era dell'intelligenza artificiale.

Dalla lettura degli atti parlamentari e delle discussioni ad esso connesse, l'articolo 8 del DDL AI pare voler introdurre una disciplina specifica per il trattamento dei dati personali nell'ambito della ricerca scientifica finalizzata allo sviluppo di sistemi di IA in campo sanitario. La norma individua come base giuridica il rilevante interesse pubblico, richiamando l'art. 9, par. 2, lett. g) del GDPR¹⁴, e trova fondamento negli articoli 32 e 33 della Costituzione¹⁵, che tutelano rispettivamente il diritto alla salute e la libertà della ricerca scientifica¹⁶. Il richiamo all'art. 33 Cost. riveste un significato particolarmente pregnante nell'economia della disciplina introdotta. Come evidenziato dalla dottrina costituzionalistica più autorevole, la libertà della ricerca scientifica proclamata dal primo comma dell'art. 33 non si esaurisce in una mera dichiarazione programmatica, ma si configura quale principio immanente al riconoscimento della scienza come "libertà stessa nella sua forma più alta"¹⁷. Tale libertà, tuttavia, non può considerarsi priva di limiti, dovendo necessariamente bilanciarsi con altri valori costituzionali che, di volta in volta, vengono posti in gioco, tutti comunque riconducibili alla dignità della persona umana.

La scelta del legislatore di ancorare la nuova disciplina all'interesse pubblico quale base giuridica autonoma per il trattamento secondario dei dati personali, vale a dire l'utilizzo per scopi diversi da quelli per i quali sono stati originariamente raccolti, rappresenta un significativo scarto rispetto al modello tradizionalmente basato sul consenso informato. Tale opzione normativa evoca l'art. 9, par. 2, lett. g) del GDPR e stimola un ripensamento dell'intero impianto garantista - consensuale che ha sinora di fatto sorretto la disciplina dei dati particolari, introducendo una logica di valorizzazione informativa che supera la dimensione meramente difensiva della privacy¹⁸. Il perimetro soggettivo della disposizione comprende enti pubblici, soggetti privati senza scopo di lucro, IRCCS e soggetti privati coinvolti in progetti congiunti, mentre le finalità spaziano dalla prevenzione e diagnosi delle malattie allo sviluppo di terapie, farmaci, dispositivi e interfacce uomo - macchina. Tale articolazione soggettiva riflette una concezione pluralista della ricerca scientifica, che riconosce la legittimità di una molteplicità di attori nella produzione di conoscenza biomedica, pur mantenendo ferme le garanzie sostanziali in ordine alla tutela dei diritti fondamentali.

¹⁴ Ai sensi dell'art. 9, par. 2, lett. g) del GDPR, il divieto di trattamento dei dati particolari può non applicarsi se «il trattamento è necessario per motivi di interesse pubblico rilevante, sulla base del diritto dell'Unione o degli Stati membri, che deve essere proporzionato all'obiettivo perseguito, rispettare l'essenza del diritto alla protezione dei dati e prevedere misure appropriate e specifiche per tutelare i diritti fondamentali e gli interessi dell'interessato».

¹⁵ Costituzione della Repubblica Italiana, art. 32, che tutela il diritto fondamentale alla salute come interesse dell'individuo e della collettività, riconoscendo limiti ai trattamenti sanitari solo nel rispetto della persona umana; art. 33, che garantisce la libertà dell'arte, della scienza e del loro insegnamento, riconoscendo autonomia alle istituzioni educative e scientifiche.

¹⁶ N.A. SMUHA (a cura di), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*, Cambridge, 2025.

¹⁷ A. SIMONCINI, *Il linguaggio dell'Intelligenza Artificiale e la tutela costituzionale dei diritti*, in *Rivista AIC*, 2025.

¹⁸ U. RUFFOLO (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milano, 2020.



L'ambito di maggiore rilevanza teorica è rappresentato dalla menzione, per la prima volta nell'ordinamento italiano, di una base giuridica specifica per le operazioni di trattamento, tra cui spicca proprio la sintetizzazione. Questa innovazione normativa riconosce formalmente lo statuto giuridico di processi computazionali che generano dati non più ancorati a referenti empirici specifici, ma costruiti algoritmicamente attraverso modelli statistici di inferenza. Si passa dalla concezione tradizionale del dato come rappresentazione della realtà empirica a una visione del dato come artefatto computazionale prodotto da processi di modellizzazione matematica: questa la cesura epistemologica di portata sistemica introdotta dalla sintetizzazione elevata a operazione giuridicamente riconosciuta.

L'art. 8 del DDL AI si colloca entro una cornice sistematica di complessità inedita, la cui analisi richiede un approccio ermeneutico che tenga conto delle profonde trasformazioni epistemologiche che l'IA introduce nel tessuto concettuale del diritto¹⁹. Questa complessità epistemologica si manifesta con particolare evidenza nelle sovrapposizioni normative, dove la coesistenza di multiple stratificazioni legislative riflette la difficoltà del legislatore nel governare fenomeni tecnologici in rapida evoluzione. L'innesto del DDL AI nel più ampio quadro normativo europeo evidenzia come la disciplina nazionale si collochi in una fase di anticipazione rispetto all'implementazione piena dell'AI Act e dello Spazio Europeo dei Dati Sanitari (EHDS). Tuttavia, questa posizione anticipatoria della legislazione italiana, pur potenzialmente virtuosa, non è priva di rischi sistemici. Per economia di trattazione e coerenza metodologica, l'analisi si concentra sulle antinomie normative interne all'ordinamento italiano, considerando che la risoluzione delle contraddizioni interne rappresenta un prerequisito necessario per un efficace coordinamento sovranazionale. La coesistenza tra l'art. 8 del DDL AI e gli artt. 110 e 110-bis del Decreto legislativo 30 giugno 2003, n. 196, comunemente noto come "Codice Privacy", configura quello che la teoria generale del diritto definisce come "eccesso di norme", richiedendo l'applicazione di criteri interpretativi per la risoluzione delle apparenti antinomie²⁰.

L'articolo 8 del DDL AI, nella sua portata innovativa, non si colloca nel vuoto normativo, ma si innesta su un impianto già articolato di disposizioni nazionali e sovranazionali che disciplinano il trattamento dei dati per finalità di ricerca scientifica, tra cui spiccano in particolare gli articoli 110 e 110-bis del Codice Privacy. La coesistenza di queste normative pone il legislatore e gli interpreti dinanzi a una complessa operazione di coordinamento sistematico, resa ancor più ardua dalla natura evolutiva dei fenomeni tecnologici cui esse si riferiscono. La crescente integrazione dell'IA nei contesti clinici e di ricerca richiede, infatti, un'analisi approfondita delle possibili sovrapposizioni, interferenze e divergenze applicative tra i diversi regimi giuridici. È proprio in questo scenario di pluralismo normativo e tensioni interpretative che si impone un'indagine mirata sulle modalità di interazione tra l'art. 8 del DDL AI e le disposizioni già vigenti del Codice Privacy, al fine di comprendere se e come possa delinearsi un modello di integrazione coerente, capace di garantire da un lato l'effettività delle tutele, dall'altro la sostenibilità giuridica dell'innovazione tecnologica.

La riforma dell'art.110 del Codice Privacy, operata dalla legge 56/2024, ha eliminato l'obbligo di consultazione preventiva del Garante, introducendo un regime di accountability rafforzata basato sul parere del comitato etico. Tale evoluzione normativa si pone in linea con i principi del GDPR, privilegiando l'autore-sponsabilizzazione dei titolari del trattamento rispetto ai controlli ex ante.

¹⁹ V. FROSINI, *Cibernetica: diritto e società*, Roma, 2023.

²⁰ G. PINO (a cura di), *Norberto Bobbio. Studi sulla teoria generale del diritto*, Torino, 2024.

Prima facie, potrebbero delinearsi almeno quattro distinte soluzioni interpretative rispetto alla coesistenza delle norme in parola. Una prima interpretazione si fonderebbe sul principio di complementarità funzionale tra le diverse discipline. L'art. 110 del Codice Privacy manterrebbe la propria applicazione generale per la ricerca scientifica tradizionale, mentre l'art. 8 del DDL AI introdurrebbe una disciplina specifica per i trattamenti finalizzati allo sviluppo di sistemi di IA²¹. Tuttavia, tale interpretazione presenterebbe criticità nella definizione dei confini applicativi, considerando la crescente integrazione delle tecnologie di IA nella ricerca biomedica contemporanea.

Un secondo scenario potrebbe ancorarsi al criterio di specialità progressiva, secondo il quale l'art. 8 del DDL AI si configurerebbe come norma speciale rispetto all'art. 110 del Codice Privacy, applicando il principio *lex specialis derogat legi generali*. L'applicazione di tale criterio favorirebbe la prevalenza dell'art. 8 del DDL AI in occasione del delinearsi di specifiche finalità di sviluppo di sistemi di IA, mantenendo ferma l'efficacia dell'art. 110 per le altre attività di ricerca. Tuttavia, questo approccio potrebbe astrattamente generare frammentazioni procedurali che potrebbero compromettere l'efficienza operativa degli enti di ricerca²².

Una terza ipotesi potrebbe essere rappresentata dal modello di concorrenza elettiva, per cui si configurerebbe la possibilità per i titolari del trattamento di scegliere tra discipline alternative in base alle specificità del progetto di ricerca. Tale interpretazione, pur offrendo flessibilità operativa, presenterebbe rischi significativi di *forum shopping* normativo che potrebbero compromettere l'uniformità applicativa delle garanzie. La scelta elettiva tra regimi normativi paralleli contrasterebbe inoltre con il principio di certezza del diritto, elemento essenziale per la tutela dell'affidamento dei cittadini e la coerenza dell'ordinamento giuridico.

Una quarta via potrebbe essere quella del modello di integrazione sistematica, secondo il quale si proporrebbe una lettura coordinata capace di valorizzare gli aspetti complementari delle diverse discipline. L'art. 110 del Codice Privacy continuerebbe ad applicarsi per le fasi di raccolta e primo trattamento dei dati, mentre l'art. 8 del DDL AI regolerebbe specificamente le operazioni di elaborazione avanzata finalizzate allo sviluppo di sistemi di IA²³. Tale interpretazione troverebbe supporto nella distinzione concettuale tra "trattamento dei dati personali per ricerca scientifica" (art. 110) e "trattamenti per la ricerca e la sperimentazione scientifica nella realizzazione di sistemi di intelligenza artificiale" (art. 8 DDL AI). La sintesi dei dati, espressamente menzionata nell'art. 8, rappresenterebbe, in tal guisa, l'elemento qualificante in grado di giustificare l'applicazione della disciplina speciale.

La complessità del quadro normativo si intensifica ulteriormente quando si considera l'art. 110-bis, comma 4 del Codice Privacy, che introduce un regime specifico per gli Istituti di Ricovero e Cura a Carattere Scientifico (IRCCS). Tale disposizione prevede che "non costituisce trattamento ulteriore da parte di terzi il trattamento dei dati personali raccolti per l'attività clinica, a fini di ricerca, da parte degli Istituti di ricovero e cura a carattere scientifico". La norma si fonda su una presunzione di compatibilità funzionale tra attività clinica e ricerca scientifica negli IRCCS, configurando una deroga specifica al principio generale del consenso informato. Come chiarito dal Garante Privacy nelle Frequently Asked Questions (FAQ) del

²¹ J.G. ALLEN, P. HUNN, *Smart Legal Contracts: Computable Law in Theory and Practice*, Oxford, 2022.

²² J. FAUCOLBRIDGE, *Trajectories of legal work in the context of machine learning AI: conceptualising mediated evolution*, in *International Journal of the Legal Profession*, 2025.

²³ A. GODBLE, *Synthetic Data for Robust AI Model Development in Regulated Enterprises*, in *IEEE*, 2025.



giugno 2024, questa norma trova applicazione per ogni tipo di ricerca medica, biomedica ed epidemiologica promossa dagli IRCCS, richiedendo unicamente lo svolgimento della Valutazione d'Impatto (DPIA) obbligatoria e la sua pubblicazione. La coesistenza di questa disciplina speciale con l'art. 8 del DDL AI solleva interrogativi interpretativi di particolare complessità, poiché coinvolge soggetti che potrebbero essere destinatari di entrambe le normative²⁴. L'applicazione del criterio di specialità suggerirebbe la prevalenza dell'art. 8 del DDL AI nei casi in cui gli IRCCS sviluppino specificamente sistemi di IA, mantenendo l'efficacia dell'art. 110-bis per le altre attività di ricerca. Tuttavia, la natura istituzionalmente integrata degli IRCCS potrebbe giustificare l'applicazione preferenziale del regime speciale dell'art. 110-bis, considerando la peculiarità organizzativa di questi enti.

L'intreccio tra l'articolo 8 del DDL AI e le disposizioni del Codice Privacy evidenzia una tensione crescente tra le esigenze di innovazione e la necessità di preservare la coerenza dell'ordinamento giuridico. In un contesto segnato da sovrapposizioni applicative, deroghe settoriali e stratificazioni regolatorie, emerge la difficoltà di governare la trasformazione tecnologica mediante i soli strumenti del diritto positivo. È in questo scenario che si affaccia, con sempre maggiore urgenza, la questione del ruolo delle fonti deontologiche come possibile leva di razionalizzazione e adattamento.

Nel contesto di crescente complessità normativa che caratterizza l'attuale panorama giuridico italiano sopra descritto, assume particolare rilevanza l'iniziativa del Garante per la protezione dei dati personali che, con il provvedimento n. 298 del 9 maggio 2024, ha avviato un processo di revisione organica delle regole deontologiche applicabili ai trattamenti di dati personali effettuati a fini statistici e di ricerca scientifica. Tale processo di elaborazione rappresenta un passaggio cruciale per il futuro della ricerca scientifica italiana, configurandosi come momento di sintesi e raccordo tra le diverse discipline che compongono l'articolato mosaico normativo in via di definizione.

Le significative difficoltà metodologiche e concettuali che stanno emergendo nel processo di riscrittura delle regole deontologiche da un lato riflettono precisamente la complessità sistemica che caratterizza l'attuale fase di transizione normativa, segnata dalla necessità di elaborare categorie giuridiche adeguate a governare fenomeni tecnologici in rapida evoluzione, dall'altro scontano la difficoltà di essere ancorate a presupposti concettuali inadeguati a rispondere efficacemente alle sfide poste dall'IA nel contesto della ricerca scientifica²⁵.

La metodologia partecipativa che caratterizza per la prima volta la disciplina delle regole deontologiche rappresenta un tentativo di superare l'inadeguatezza concettuale attraverso un modello di co-regolazione che coinvolge stakeholder, comunità scientifica e società civile. Tuttavia, questo approccio, pur innovativo, rischia di risultare insufficiente se non accompagnato da una radicale riconsiderazione dei fondamenti epistemologici che sottendono la regolamentazione dell'IA. La sfida fondamentale consiste nel superare quella che potremmo definire una "dissonanza epistemica" tra i presupposti concettuali della normativa tradizionale e la natura intrinsecamente probabilistica, adattiva e non-deterministica dei sistemi di IA. Questa dissonanza è evidente nell'ambito della ricerca scientifica, dove l'IA non si limita a replicare metodologie consolidate, ma introduce modalità di generazione della conoscenza radicalmente nuove, basate su processi di inferenza statistica e apprendimento automatico che sfuggono alle categorie

²⁴ A.A.V.V. *Il futuro regolatorio dei dati sintetici*, in *Lumsa Better Regulation*, 2024.

²⁵ R. WHALEN (a cura di), *Computational Legal Studies: The Promise and Challenge of Data-Driven Research*, Cheltenham, 2020.

tradizionali di validazione scientifica²⁶. Le regole deontologiche, in questo contesto di transizione paradigmatica, sono chiamate a svolgere una funzione che va oltre la mera complementarità rispetto alla disciplina legislativa, configurandosi come autentico laboratorio per lo sviluppo di nuove categorie concettuali capaci di cogliere la specificità dei sistemi di IA. La loro riscrittura non rappresenta semplicemente un aggiornamento tecnico, ma un'operazione di profonda riconsiderazione dei fondamenti stessi del rapporto tra diritto e tecnologia, tra norma e prassi, tra principio e applicazione²⁷.

L'analisi comparativa degli scenari interpretativi evidenzia come il modello di integrazione sistematica, sebbene apparentemente possa rappresentare la soluzione più coerente con i principi di accountability e proporzionalità che caratterizzano l'evoluzione del diritto della protezione dei dati, tuttavia non è immune da critiche o mancanze. Appare utile infatti l'identificazione di un approccio che consenta di valorizzare le specificità dell'IA senza compromettere la coerenza sistemica dell'ordinamento, utilizzando le regole deontologiche come strumento di raccordo e armonizzazione tra le diverse discipline.

La specificità epistemologica dell'IA, caratterizzata da processi di sintetizzazione e modellizzazione algoritmica, giustifica l'introduzione di strumenti normativi ad hoc che tengano conto delle peculiarità tecniche e metodologiche del settore. L'art. 8 del DDL AI, con il riconoscimento esplicito della sintetizzazione come operazione giuridicamente rilevante, introduce per la prima volta nell'ordinamento italiano una disciplina organica per i dati sintetici. L'evoluzione verso regimi di accountability rafforzata, inaugurata dalla riforma dell'art. 110 del Codice Privacy, rappresenta un trend normativo coerente con i principi europei, e l'art. 8 del DDL AI si inserisce in questo quadro evolutivo, offrendo strumenti specifici per la governance dell'IA senza compromettere le garanzie sostanziali di protezione dei dati personali.

In questo quadro, l'evoluzione normativa tracciata dall'art. 8 del DDL AI non esaurisce il ventaglio delle innovazioni legislative rilevanti. Parallelamente al riconoscimento della sintetizzazione come operazione giuridicamente rilevante, ulteriori disposizioni contenute nel DDL aprono scenari ancora più dinamici sul piano regolatorio, introducendo strumenti sperimentali volti a coniugare innovazione, flessibilità e supervisione istituzionale. Tra queste, merita una riflessione specifica l'art. 9, che richiama esplicitamente l'istituzione di spazi regolatori dedicati alla sperimentazione in ambito sanitario, evocando il modello dei regulatory sandbox.

L'articolo 9 del DDL AI, inserito nel corso dell'esame parlamentare, e che rimette a un decreto del Ministro della Salute la disciplina del trattamento dei dati personali, anche particolari, per finalità di ricerca e sperimentazione tramite sistemi di IA, inclusa la costituzione di spazi speciali di sperimentazione evoca il concetto di regulatory sandbox, strumento di crescente rilevanza nel panorama della regolamentazione dell'innovazione tecnologica. I regulatory sandbox, come definiti dall'art. 3, n. 55 dell'AI Act, si configurano come "un quadro controllato istituito da un'autorità competente che offre ai fornitori o potenziali fornitori di sistemi di IA la possibilità di sviluppare, addestrare, convalidare e provare, se del caso in condizioni reali, un sistema di IA innovativo, conformemente a un piano dello spazio di sperimentazione per un periodo di tempo limitato sotto supervisione regolamentare". Essi hanno la funzione primaria di consentire lo sviluppo dei sistemi di IA, in particolare sistemi potenzialmente ad alto rischio, all'interno di un ambiente controllato.

²⁶ B.K. CHAU, M.A. Livermore, *Computational Legal Studies Comes of Age*, in *European Journal of Empirical Legal Studies*, Vol. 1, n. 1, 2024.

²⁷ L. LESSIG, *Code: And Other Laws of Cyberspace. Version 2.0*, New York, 2006.



Secondo l'art. 57 dell'AI Act, ogni Stato membro deve stabilire almeno un regulatory sandbox per l'IA a livello nazionale entro il 2 agosto 2026²⁸. Questi ambienti controllati migliorano la certezza giuridica, supportano la compliance, consentono il trattamento di dati personali e facilitano l'accesso al mercato per PMI e startup. Tuttavia, la configurazione di questi spazi di sperimentazione richiede un'attenta valutazione della compatibilità con l'art. 57 dell'AI Act, che disciplina i regulatory sandbox a livello europeo, particolarmente nella parte implementativa affidata al Ministero della Salute.

L'inserimento dei sandbox regolatori nel disegno di legge, lungi dall'essere una mera innovazione procedurale, si inserisce in un più ampio movimento di trasformazione del diritto, che tende a ridefinire in profondità i presupposti epistemici e normativi del trattamento dei dati a fini di ricerca. È proprio da questa prospettiva sistemica che occorre ora guardare all'art. 8 del DDL AI: non come norma isolata, ma come segnale di una nuova razionalità giuridica, capace di riorganizzare l'intero impianto delle condizioni di liceità nel contesto dell'innovazione medico-scientifica.

L'articolato confronto interpretativo sull'art. 8 del DDL n. 1146/2024, condotto nei paragrafi precedenti, ha evidenziato come l'ordinamento italiano stia progressivamente elaborando una disciplina capace di riconoscere specificità tecniche della ricerca medica fondata sull'IA. La menzione espressa della sintetizzazione, il superamento del paradigma consensuale e l'apertura ai regulatory sandbox indicano una mutazione di fondo del quadro regolatorio, orientata a rendere possibile e legittimo un uso innovativo dei dati a supporto della conoscenza scientifica.

Per comprendere appieno la portata di questa transizione, è necessario spostare lo sguardo dal singolo provvedimento normativo alla dinamica più ampia che investe l'intero sistema delle condizioni di liceità del trattamento per finalità di ricerca nel panorama nazionale.

La norma in esame si colloca infatti in un contesto di evoluzione strutturale che interroga le basi giuridiche tradizionali, e in particolare, la tenuta del consenso come fondamento esclusivo del trattamento dei dati sanitari. Le recenti riforme del Codice Privacy, l'implementazione delle Privacy-enhancing Technologies (PET) e la crescente attenzione alle logiche di interesse pubblico mostrano l'emergere di un nuovo equilibrio tra tutela dei diritti e abilitazione della ricerca. Il paragrafo successivo si soffermerà proprio su questa traiettoria evolutiva, esaminando come l'ordinamento stia ridefinendo i presupposti normativi per il trattamento dei dati personali in ambito medico scientifico, nella prospettiva di un sistema giuridico orientato all'evidenza e in grado di accompagnare l'innovazione senza rinunciare alle garanzie fondamentali.

2.3. L'evoluzione delle condizioni di liceità per il trattamento di dati personali per scopi di ricerca medica

L'introduzione, da parte dell'art. 8 del DDL n. 1146/2024, di una base normativa specifica per il trattamento dei dati personali, anche appartenenti a categorie particolari, per finalità di ricerca e sperimentazione scientifica mediante sistemi di IA, evidenzia una presa di posizione circa l'inadeguatezza, in molteplici contesti applicativi, del consenso quale presupposto giuridico sufficiente. L'esplicito riconoscimento dell'uso secondario dei dati e della sintetizzazione tra le operazioni lecite costituisce una risposta normativa alle difficoltà in merito alla sostenibilità del consenso in ambito scientifico.

Tale impostazione si inserisce in un contesto più ampio, in cui da tempo si sollevano dubbi circa l'effettiva idoneità del consenso quale base giuridica del trattamento dei dati sanitari per scopi di ricerca medica.

²⁸ *AI Regulatory Sandbox Approaches: EU Member State Overview*, 2025.

Già nel 2019, il Comitato europeo per la protezione dei dati metteva in guardia i titolari sull'effettiva idoneità del consenso, ai sensi degli artt. 6, par. 1 lett. a) e 9, par. 2, lett. a) GDPR, come base giuridica del trattamento dei dati sulla salute nelle sperimentazioni cliniche ai sensi del Regolamento (UE) n. 536/2014; ma medesime considerazioni possono valere per gli studi clinici retrospettivi che non prevedono la sperimentazione di un farmaco²⁹. Ciò, in quanto la condizione di vulnerabilità degli interessati può comprometterne il requisito di libertà, mentre quello della revocabilità può incidere sul progredire della ricerca. La caducità del consenso in tale contesto può essere inoltre imputata alla materiale impossibilità di acquisirlo; si aggiunga, infine, che esso non costituisce una valida condizione di liceità per il trattamento di dati raccolti e trattati in esecuzione di un compito di interesse pubblico o di un obbligo legale (cons. 43 GDPR). Alla luce delle questioni emerse, risulta opportuno soffermarsi sulle basi giuridiche che, in alternativa al consenso, possono fondare il trattamento dei dati sanitari per finalità di ricerca medica, nel rispetto dei principi delineati dal GDPR e tenendo conto delle esigenze della comunità scientifica. In tale prospettiva, si procede a una ricognizione delle disposizioni normative rilevanti, a livello europeo e nazionale, con specifico riferimento all'art. 9, par. 2, lett. g) e j) del GDPR e ai criteri di proporzionalità e prevedibilità che ne guidano l'applicazione.

Ciò stante, avendo a mente le dieci esenzioni dal divieto di trattare le particolari categorie di dati, di cui all'art. 9, par. 2 GDPR, nella concreta impossibilità di potersi utilizzare il consenso, l'invito di trovare ulteriori e diverse basi giuridiche non può che tradursi nella ricerca di una normativa dell'ordinamento nazionale o unionale, proporzionata alla finalità perseguita, coerente con l'essenza del diritto alla protezione dei dati e che preveda misure appropriate e specifiche per tutelare i diritti fondamentali e gli interessi degli interessati, anche in conformità all'articolo 89 del GDPR (art. 9, par. 2, lett. g) e j) GDPR). I ricercatori alternativamete dovrebbero utilizzare dati che non siano di carattere personale, nella misura in cui essi siano comunque utili ai loro scopi.

Sotto un primo profilo, vanno quindi tenuti a mente il monito di cui al cons. 41 GDPR, secondo il quale la misura legislativa sulla quale si fonda un trattamento dovrebbe essere *"chiara e precisa, e la sua applicazione prevedibile, per le persone che vi sono sottoposte"*, e la previsione di cui all'art. 36, par. 4 GDPR, secondo cui *"Gli Stati membri consultano l'autorità di controllo durante l'elaborazione di una proposta di atto legislativo [...] o di misura regolamentare basata su detto atto legislativo relativamente al trattamento"*³⁰. Rilevano, inoltre, le indicazioni pratiche del Garante europeo per la protezione dei dati (EDPS) per la verifica della necessità a proporzionalità dei trattamenti, quelle rivolte alle istituzioni comunitarie sui principali elementi da considerare nella predisposizione di proposte legislative e gli altri atti normativi che implicano il trattamento dei dati personali (applicabili per analogia anche nei contesti nazionali), dove l'EDPS indica, in particolare, che *"greater the interference entailed by the proposed act, the most robust and detailed the rules and the safeguards should be"*.

²⁹ Comitato europeo per la protezione dei dati: *Parere 3/2019 relativo alle domande e risposte sull'interazione tra il regolamento sulla sperimentazione clinica e il regolamento generale sulla protezione dei dati*, 23 gennaio 2019; Linee guida 5/2020 sul consenso ai sensi del regolamento (UE) 2016/679, adottate del 4 maggio 2020.

³⁰ Garante europeo per la protezione dei dati: *Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit*, 11 aprile 2017; *Guidelines on assessing the proportionality of measures that limit the fundamental rights to privacy and to the protection of personal data*, 19 dicembre 2019; *Guidance for co-legislators on key elements of legislative Proposals*, 7 maggio 2025.



Tenendo a mente tali indicazioni, si ritiene inoltre che il quadro giuridico di riferimento debba ora essere duttile, senza essere incerto, prevedendo normative di rango primario che contengano gli elementi essenziali del trattamento e che si sviluppino poi attraverso atti attuativi idonei ad essere, se del caso, facilmente aggiornati, tenuto conto del costante progredire della tecnologia e delle conseguenti e nuove esigenze della comunità scientifica.

Occorre quindi chiedersi se l'evoluzione del panorama normativo (concentrandosi in questa sede solo su quello nazionale) si stia dirigendo in questo senso. In tale direzione, si rende utile ricostruire l'evoluzione recente del quadro normativo nazionale, al fine di delineare gli strumenti giuridici attualmente disponibili per abilitare il trattamento dei dati sanitari a fini di ricerca, in coerenza con i principi europei di necessità, proporzionalità e tutela dei diritti.

A tale riguardo, merita quindi menzionare la modifica, apportata dal d.l. 2 marzo 2024, n. 19, all'art. 110 del Codice della privacy, nella parte in cui introduce una condizione di liceità "alternativa" al consenso, quando risulti impossibile acquisirlo. Abrogando l'obbligo di consultazione preventiva ex art. 36 GDPR, la novella ha affidato al Garante il compito di individuare specifiche garanzie, concretamente definite nella deliberazione del 9 maggio 2024³¹.

L'ordinamento italiano dispone poi di uno specifico corpo normativo sui cosiddetti sistemi di sorveglianza e registri di patologia, disciplinando, attraverso decreti ministeriali, adottati previo parere del Garante, condizioni e limiti per il trattamento dei dati sulla salute ivi raccolti anche per fini di ricerca medica (art. 12, commi da 10 a 14 del d.l. 18 ottobre 2012, n. 179 e ss.ii.mm; d.P.C.M. 3 marzo 2017, con. 157 GDPR). Anche il Fascicolo Sanitario Elettronico (FSE) costituisce uno strumento astrattamente utile anche a fini di ricerca medica (art. 12, comma 2, lett. b) del d.l. 179, art. 27 d.m. 2 settembre 2023).

Tuttavia, è nel d.m. del 31 dicembre 2024 istitutivo dell'EDS (Ecosistema di dati sanitari), previsto all'art. 12, comma 15-*quater* dal d.l. 179 del 2012, che si rinviene una prima indicazione normativa per l'uso dei dati e documenti del cosiddetto FSE 2.0 per fini di ricerca, laddove è previsto che esso renda disponibili servizi di estrazione dei dati anonimizzati per scopi di ricerca medica, rinviando ad ulteriori decreti per l'uso dei predetti dati in forma pseudonimizzata (art. 17).

Si richiama l'attenzione sui commi 1-bis e 1-ter dell'art. 2-sexies del Codice Privacy, i quali prevedono che, con decreti del Ministero della Salute adottati previo parere del Garante, sia:

- disposta l'interconnessione a livello nazionale dei sistemi informativi contenenti dati pseudonimizzati relativi alla salute, detenuti dal Ministero della salute, dall'Istituto Superiore di Sanità, dall'Agenzia nazionale per i servizi sanitari regionali, dall'AIFA, dall'Istituto Nazionale salute, Migrazioni e Povertà, nonché dalle Regioni e dalle Province autonome;
- disciplinato un ambiente di trattamento sicuro, all'interno del quale tali dati (pseudonimizzati o anonimizzati, inclusi quelli del Fascicolo Sanitario Elettronico e di altre PA) possano essere messi a disposizione per il perseguimento delle finalità istituzionali dei rispettivi enti.

³¹ Garante per la protezione dei dati personali: *Parere sullo schema di decreto del Ministero della salute, da adottare assieme al Ministro delegato per l'innovazione tecnologica e la transizione digitale, di concerto con il Ministro dell'economia e delle finanze, sul Fascicolo Sanitario Elettronico (FSE)*, 8 giugno 2023 (doc. web 9900433); *Parere sullo schema di decreto del Ministero della salute sull'Ecosistema Dati Sanitari (EDS), ai sensi dell'art. 12, comma 15-*quater*, del decreto-legge 18 ottobre 2012, n. 179*, 26 settembre 2024 (doc. web 10062302); *Deliberazione di promulgamento delle Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica ai sensi degli artt. 2-*quater* e 106 del Codice*, 9 maggio 2024 (doc. web 10016146)

Laddove, quindi, tra le finalità istituzionali degli enti coinvolti nella richiamata interconnessione vi sia altresì quella di ricerca scientifica, essi potranno beneficiare di tali norme come base normativa per il trattamento dei dati in possesso delle altre amministrazioni considerate, nei limiti e alle condizioni che verranno definite nella normativa di attuazione.

Merita infine soffermarsi nuovamente sul Disegno di legge n. 1146, oggetto di analisi del precedente paragrafo. Come già visto, la norma riconosce alla ricerca e alla sperimentazione scientifica la natura di interesse pubblico rilevante. Per tali fini “l’uso secondario” è considerato sempre ammesso, senza ulteriore consenso, potendo essere avviati decorsi trenta giorni dalla comunicazione, salvo diverso avviso dell’Autorità (art. 8, commi 1, 2 e 5).

La norma dichiara inoltre sempre consentita, in particolare ai fini sopra descritti, la “*sintetizzazione dei dati personali*” anche inerenti alle particolari categorie, affidando ad apposite linee guida dell’AGENAS, previo parere del Garante, la definizione delle procedure in particolare di sintetizzazione dei dati personali.

È quindi la norma stessa che si occupa di fornire alla comunità scientifica una possibile alternativa all’uso di dati di carattere personale facendo leva sulle cosiddette *Privacy-enhancing technologies* (PET), quelle cioè atte ad operare la sintetizzazione dei dati.

In conclusione, l’ordinamento italiano, ponendosi nel solco delle lettere j) o g) dell’art. 9, par. 2 del GDPR e affidando la normativa di attuazione a strumenti differenti (provvedimenti del Garante, decreti ministeriali, linee guida), sta evolvendo nel senso sopra delineato attraverso norme maggiormente idonee per forma e contenuto a favorire l’effettivo e pieno svolgimento dell’attività di ricerca.

Quest’ultima non può più trovare soddisfazione nei (soli) strumenti del passato, quale ad esempio la raccolta diretta dei dati presso gli interessati, ma anche attraverso l’uso di nuove tecnologie, assicurando al contempo adeguata tutela ai diritti e alle libertà fondamentali degli interessati, in particolare attraverso misure di pseudonimizzazione, il ricorso alle PET e grazie anche al coinvolgimento dell’Autorità Garante nel processo normativo.

2.4. Evidence by design: la funzione orientante del Regolamento HTA nel governo della sperimentazione.

Nel solco delle trasformazioni normative delineate a livello nazionale di cui sopra, dalla riformulazione dell’art. 110 del Codice Privacy alla disciplina innovativa introdotta dal disegno di legge n. 1146/2024, si consolida una traiettoria evolutiva volta a superare la rigidità del modello fondato sul consenso individuale, in favore di un impianto più duttile, capace di coniugare esigenze di tutela e istanze di progresso scientifico. Questa transizione normativa, ancora legata a categorie giuridiche in fase di ridefinizione, si colloca all’interno di un più ampio processo di razionalizzazione che interessa l’intero spazio giuridico europeo in materia sanitaria. Nell’attuale fase di consolidamento del diritto dell’Unione europea in materia sanitaria, il Regolamento (UE) 2021/2282 sulla valutazione delle tecnologie sanitarie, entrato in vigore il 12 gennaio 2025, si configura come un dispositivo normativo emblematico della nuova razionalità regolatoria che permea le dinamiche del mercato unico.

Esso incarna il tentativo di coniugare le esigenze di efficienza economica, competitività industriale e sostenibilità dei sistemi sanitari con la tutela sostanziale del diritto alla salute, iscritto nella matrice



costituzionale dell'ordinamento europeo. Il combinato disposto degli articoli 114 e 168 TFUE³² ne costituisce l'impalcatura giuridica³³, riflettendo una tensione strutturale tra uniformazione tecnica e rispetto delle prerogative sovrane degli Stati membri. La governance che ne discende è di tipo cooperativo, fondata su una convergenza metodologica piuttosto che su una armonizzazione coattiva, ove l'*evidence-based regulation* diventa principio ispiratore e strumento di legittimazione³⁴. Lungi dal ridursi a una mera procedura tecnico-amministrativa, il Regolamento si presenta quale snodo concettuale di una più ampia trasformazione della funzione valutativa in ambito sanitario: la valutazione della tecnologia sanitaria (Health Technology Assessment – HTA)³⁵ non è più intesa come momento separato e successivo all'innovazione tecnologica, ma come principio immanente alla sua stessa progettazione. In tale prospettiva, la distinzione tra valutazione *ex ante* ed *ex post* tende a sfumare, lasciando emergere una nuova grammatica normativa in cui i confini tra produzione del sapere scientifico e costruzione giuridica della decisione si intrecciano inestricabilmente. L'impatto sistemico di tale approccio investe in primo luogo la ricerca clinica, chiamata a riformulare le proprie architetture metodologiche in funzione dei criteri di trasparenza, riproducibilità e comparabilità che presiedono all'HTA. Le finalità perseguite dal Regolamento, espressamente enunciate sin dall'articolo 1, riflettono l'ambizione di rafforzare il livello di protezione della salute pubblica, promuovere un uso efficiente delle risorse e garantire un accesso tempestivo ed equo alle tecnologie sanitarie innovative³⁶. A esse si affianca la volontà di ridurre la frammentazione valutativa tra gli Stati membri, favorendo processi decisionali condivisi attraverso meccanismi di valutazione clinica congiunta, nonché la costruzione di una base comune di evidenze scientifiche³⁷. Si tratta di una finalità che, pur collocandosi formalmente nell'ambito della valutazione, finisce per irradiare i suoi effetti sull'intero ciclo di vita della tecnologia sanitaria, incidendo anche sulle scelte metodologiche che presiedono alla sua sperimentazione clinica. I principi ispiratori che strutturano l'intervento regolatorio, quali l'indipendenza, l'imparzialità, la trasparenza, il rispetto delle competenze nazionali e la proporzionalità, delineano una forma inedita di normazione integrata, dove il diritto si fa facilitatore di processi valutativi sovranazionali

³² Il combinato disposto degli articoli 114 e 168 TFUE costituisce l'architrave giuridica su cui si fonda il Regolamento (UE) 2021/2282, esprimendo la tensione strutturale tra le logiche di integrazione funzionale del mercato interno e la tutela sostanziale della salute pubblica. L'art. 114, nella sua funzione armonizzatrice, legittima interventi normativi volti a garantire l'unità del mercato, anche in ambiti sensibili come quello sanitario, a condizione che sia assicurato un elevato livello di protezione. È a partire da tale base che l'Unione ha sviluppato strumenti di regolazione condivisa e dispositivi di *soft law* finalizzati a superare le disomogeneità nazionali. L'art. 168, specularmente, opera come clausola di salvaguardia, imponendo che ogni politica unionale integri il principio di tutela della salute e ribadendo la competenza degli Stati membri nell'organizzazione dei servizi sanitari. La dialettica tra queste due disposizioni alimenta una forma di governance cooperativa, in cui esigenze di efficienza.

³³ A. DE RUIJTER, *EU Health Law & Policy: The Expansion of EU Power in Public Health and Health Care*, Oxford, 2019.

³⁴ Regolamento (UE) 2021/2282, considerando 1, 2 e 10.

³⁵ Nel Regolamento (UE) 2021/2282, il termine «valutazione» è definito all'articolo 2, paragrafo 1, lettera a) come segue: «valutazione della tecnologia sanitaria o «HTA»: un processo multidisciplinare che sintetizza informazioni sugli aspetti medici, organizzativi, economici, sociali, etici e giuridici relativi all'uso di una tecnologia sanitaria in maniera sistematica, trasparente, imparziale e solida. Il suo scopo è informare l'elaborazione di politiche sanitarie sicure, efficaci, centrate sul paziente e che offrano valore aggiunto.»

Tale definizione evidenzia la natura composita e scientificamente strutturata della HTA, intesa non come una mera analisi tecnica o economica, ma come un'attività che integra la dimensione clinica, etica, e giuridica per supportare le decisioni pubbliche in sanità.

³⁶ Art. 1, par. 1, Regolamento (UE) 2021/2282.

³⁷ Considerando 11 e 14, Regolamento (UE) 2021/2282.

senza però giungere a comprometterne l'autonomia decisionale³⁸. È proprio in questa tensione tra uniformità metodologica e pluralismo decisionale che si annidano alcune delle principali criticità interpretative: da un lato, si promuove una convergenza tecnica sulla metodologia valutativa, dall'altro si riafferma il carattere non vincolante degli esiti valutativi, che gli Stati sono tenuti a considerare ma non ad assumere³⁹. Tale ambivalenza produce un quadro giuridico fluido, che sfida la tradizionale dicotomia tra hard law e soft law, collocandosi in una zona grigia che impone una rinnovata riflessione sul principio di effettività nella regolazione dell'innovazione⁴⁰.

Proprio in questa cornice si fa strada l'istanza di ripensare le forme canoniche della sperimentazione clinica, a partire dalla struttura dei bracci di controllo. Non si tratta unicamente di adeguare gli studi a una nuova regolazione, ma di riconoscere che la stessa costruzione dell'evidenza deve ormai avvenire in funzione dei criteri di valutazione comparativa⁴¹. L'impiego di strumenti computazionali avanzati, tra cui l'IA e i dati sintetici, non è più una variabile meramente tecnologica, bensì una componente strutturale della nuova epistemologia della ricerca⁴². Alcuni studi iniziano a essere progettati secondo logiche che incorporano sin dall'inizio le finalità dell'HTA, prospettando una prassi scientifica che si orienta già *ex ante* all'idoneità valutativa del dato prodotto.

L'introduzione dei dati sintetici nei bracci di controllo rappresenta, in questo senso, uno snodo cruciale. Da un lato, tali dati permettono di ridurre il numero di pazienti esposti a trattamenti subottimali o placebo, impattando positivamente sulle implicazioni etiche della sperimentazione; dall'altro, pongono interrogativi complessi circa la loro legittimità regolatoria e la loro capacità di produrre evidenze scientificamente robuste che dovranno essere validate dagli enti regolatori. Il Regolamento, pur non affrontando direttamente la questione, ne predispone i presupposti normativi: ciò che rileva non è l'origine del dato, bensì la sua qualità, trasparenza, replicabilità e valore comparativo. Se debitamente validata, la sintetizzazione del braccio di controllo potrebbe rappresentare una forma di pieno adempimento delle finalità regolatorie, e non una loro elusione. In questa prospettiva la trasformazione in atto nella sperimentazione clinica non è semplicemente un'evoluzione tecnica, ma un mutamento paradigmatico che interroga il diritto nella sua funzione ordinante e orientante. Il Regolamento HTA, nella sua architettura funzionale, si propone come catalizzatore di tale mutamento, sollecitando una riformulazione della ricerca clinica che non sia più mera verifica empirica, ma costruzione teleologicamente orientata dell'evidenza. I bracci di controllo, lungi dall'essere uno spazio neutro del disegno sperimentale, diventano il luogo in cui si misurano le sfide dell'innovazione, dell'equità e della sostenibilità dei sistemi sanitari.

L'integrazione dei dati sintetici, previa validazione metodologica che tenga conto di una valutazione clinica rigorosa, si presenta come il passaggio necessario da una logica di registrazione passiva del reale a una

³⁸ Art. 5 e considerando 44, Regolamento (UE) 2021/2282.

³⁹ Art. 8, par. 5, Regolamento (UE) 2021/2282.

⁴⁰ F. PASQUALE, *The Black Box Society*, Cambridge, 2015.

⁴¹ La costruzione dell'evidenza scientifica non può più essere pensata come un processo autonomo, indipendente, guidato solo da criteri clinici o sperimentali interni ma deve essere progettata fin dall'inizio tenendo conto dei criteri propri della valutazione comparativa previsti dal Regolamento: solidità, trasparenza, replicabilità e valore nel confronto tra tecnologie. Il metodo con cui si produce la conoscenza deve già essere conforme al metodo con cui quella conoscenza verrà giudicata utile per decisioni sanitarie pubbliche.

⁴² Trattasi di impostazione già sperimentata in alcuni protocolli clinici: si veda progetto VEN-DEC al primo capitolo della seconda parte del lavoro.

logica attiva di generazione “regolata” dell'evidenza. È su questo terreno che si gioca oggi la capacità del diritto pubblico di accompagnare, e non ostacolare, l'innovazione, elaborando quadri normativi che non si limitino a contenere la novità, ma che sappiano, con misura e responsabilità, valorizzarne il potenziale trasformativo nella cornice della legalità costituzionale ed europea.

2.5. Un primo bilancio

Il percorso tracciato dai paragrafi che compongono il presente capitolo evidenzia la necessità di un ripensamento strutturale del rapporto tra diritto e conoscenza nell'era dell'IA.

La nozione tradizionale di dato personale, legata alla referenzialità empirica del soggetto, si mostra inadeguata di fronte a nuove e inedite forme identitarie generate da processi computazionali. Il diritto è chiamato a confrontarsi non soltanto con nuove tecnologie, ma con una trasformazione del proprio linguaggio: ciò che entra in crisi non è l'oggetto della regolazione, ma il linguaggio stesso con cui il diritto tradizionalmente lo ha nominato e ordinato. In tale cornice la sintetizzazione dei dati è sintomo della trasformazione in atto: non mera tecnologia, ma pratica che interroga le categorie classiche del diritto. Il DDL AI testimonia un primo tentativo di normare questa transizione, introducendo nel tessuto giuridico nazionale istituti che, seppur ancora vincolati a categorie tradizionali, manifestano la consapevolezza della necessità di un diritto capace di orientarsi nella complessità introdotta dall'IA. Nella stessa direzione, la riforma dell'art. 110 del Codice Privacy e l'evoluzione verso una governance interconnessa segnano la mutazione di un diritto più modulare e adattivo. Questa trasformazione assume portata sistemica con il Regolamento (UE) 2021/2282 sull'HTA, che sposta l'asse della valutazione dall'ex post all'evidence by design, facendo del diritto un co-artefice dell'infrastruttura conoscitiva della ricerca.

In tale quadro, sarà determinante il coordinamento tra le fonti normative e il ruolo che AGENAS sarà chiamata a svolgere secondo quanto prescritto dal comma 4 dell'art. 8, in concerto con il Garante per la protezione dei dati personali, rappresentando un elemento non meramente accessorio ma strutturalmente necessario per garantire l'effettività del sistema di tutele e la coerenza dell'ordinamento nel suo complesso. Parimenti fondamentale si configura l'applicazione coordinata del sistema normativo attraverso puntuali chiarimenti interpretativi che dovranno essere emanati dal Garante Privacy, poiché la pubblicazione di FAQ specifiche sui rapporti tra le diverse discipline giuridiche coinvolte non costituisce semplicemente un'opzione opportuna, ma si rivela assolutamente necessaria per assicurare certezza applicativa e uniformità delle tutele su tutto il territorio nazionale.

In definitiva, il DDL AI, con particolare riferimento alla disciplina contenuta negli articoli 8 e 9, si configura quale tentativo sistematico e metodologicamente ambizioso di governare la transizione verso paradigmi inediti di ricerca scientifica, offrendo strumenti normativi innovativi che, se correttamente interpretati e coerentemente applicati, potranno contribuire in modo significativo allo sviluppo responsabile dell'IA nel settore sanitario. Trattasi di una sfida ermeneutica che consiste nel valorizzare le potenzialità di questa disciplina normativa superando le antinomie sistemiche di cui sopra mediante un'interpretazione evolutiva che tenga conto e delle specificità tecnologiche emergenti e dei principi fondamentali dell'ordinamento giuridico, presupponendo necessariamente una collaborazione istituzionale matura e metodologicamente rigorosa tra le diverse Autorità e Agenzie coinvolte. Tale collaborazione dovrà essere sviluppata attraverso un dialogo interistituzionale che, pur nel rispetto delle specifiche competenze e delle diverse

funzioni, si caratterizzi per la reciprocità, la trasparenza e l'orientamento collaborativo in favore del perseguimento dell'interesse pubblico sotteso all'innovazione responsabile nel settore sanitario.

3. I dati sintetici come paradigma emergente per la valutazione delle tecnologie sanitarie (HTA): dalla protezione alla valorizzazione del dato

3.1. Il contributo dei dati sintetici alla disponibilità e interoperabilità dell'evidenza

I dati, come già emerso chiaramente nel precedente capitolo, rivestono un ruolo centrale nell'innovazione in ambito sanitario, rappresentando il fattore determinante in processi eterogenei come le sperimentazioni cliniche, lo sviluppo di dispositivi medici, l'innovazione nelle procedure terapeutiche, nonché la creazione di sistemi di IA applicati alla medicina e alla sanità digitale. Il già citato Regolamento dell'Unione Europea sulla valutazione delle tecnologie sanitarie, mira a razionalizzare i processi valutativi di tali innovazioni, favorendo “[...] un accesso più rapido e diffuso a prodotti innovativi, nuovi ed efficaci per i pazienti”⁴³. Tra gli obiettivi principali figura l'istituzione di procedure accelerate per la valutazione HTA, inizialmente riservate ai farmaci oncologici o ai medicinali per terapie avanzate, con una successiva estensione prevista per il 2026, il 2028 e il 2030.

Per un'attuazione efficace di tale regolamento, è fondamentale disporre di evidenze complete e di alta qualità, in modo da garantire una valutazione accurata delle nuove tecnologie sanitarie sotto il profilo della sicurezza, dell'efficacia e della sostenibilità economica. Il passaggio generalizzato alla gestione elettronica dei dati ha prodotto un'abbondanza di informazioni archiviate⁴⁴; tuttavia, le metodologie tradizionali di raccolta dati continuano a presentare ostacoli rilevanti alla disponibilità e interoperabilità dell'evidenza. Tali ostacoli, oltre a problematiche logistiche e operative, includono questioni etiche e giuridiche legate alla protezione dei dati personali.

Per quanto riguarda la privacy, le normative e i protocolli in materia determinano cicli di ricerca lunghi e costosi, a causa della necessità di gestire richieste di accesso ai dati⁴⁵. Tra queste rientra l'ottenimento delle autorizzazioni all'accesso, un processo oneroso e dispendioso in termini di tempo. Le tecnologie di de-identificazione vengono talvolta considerate un'alternativa a tali autorizzazioni; tuttavia, il modello “consenso o anonimizzazione” si è rivelato sistematicamente inadeguato, in quanto le tecnologie convenzionali di de-identificazione non garantiscono un livello sufficiente di protezione dei dati⁴⁶. Oltre a questa insufficienza, tali tecnologie risultano anche ad alta intensità di risorse.

⁴³ https://ec.europa.eu/commission/presscorner/detail/en/ip_25_226 (ultima consultazione 21/06/2025).

⁴⁴ S.A. SENTHILKUMAR, B.K. RAI, A.A. MESHRAM, A. GUNASEKARAN, S. CHANDRAKUMARMANGALAM, *Big data in healthcare management: a review of literature*, in *American Journal of Theoretical and Applied Business*, 4(2), 2018, 57-69.

⁴⁵ A tal proposito si veda R.B. NESS, JOINT POLICY COMMITTEE, SOCIETIES OF EPIDEMIOLOGY, *Influence of the HIPAA Privacy Rule on health research*, in *JAMA*, 298(18), 2007, 2164-2170. doi: 10.1001/jama.298.18.2164. Si veda anche J. LANE, C. SCHUR *Balancing access to health data and privacy: a review of the issues and approaches for the future*, in *Health Services Research*, 45(5 pt. 2), 2010, 1456-1467.

⁴⁶ A tal proposito si veda M. MOSTERT, A.L. BREDENOORD, M.C. BIESAART, J.J. VAN DELDEN, *Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach*, in *European Journal of Human Genetics*, 24(7), 2016, 956-960. Si veda anche B.D. MITTELSTADT, L. FORIDI (a cura di), *The Ethics of Biomedical Big Data*, Cham, 2016. Si veda anche L. SEVENS, G. LAURIE, *The Administrative Data Research Centre Scotland: a scoping report on the legal*.

Le criticità logistiche e pratiche incidono ulteriormente sulla disponibilità dei Real-World Data (RWD)⁴⁷. Ad esempio, le informazioni raccolte, come quelle contenute nelle cartelle cliniche elettroniche (electronic health records, EHR), risultano spesso incomplete⁴⁸. Possono, ad esempio, presentare valori mancanti, che vanno da attributi non registrati relativi a specifici pazienti fino alla totale assenza di variabili confondenti rilevanti. Le banche dati mediche, inoltre, sono soggette a errori, come documentato in un contributo pionieristico di Komaroff del 1979⁴⁹, da allora considerato una conoscenza acquisita nel settore⁵⁰. Tali errori possono derivare da inserimenti errati⁵¹, interpretazioni fuorvianti o altre cause, con impatti potenzialmente significativi sui risultati degli studi.

I dati sanitari, inoltre, sono frequentemente affetti da bias⁵². Tali distorsioni possono assumere forme statistiche tradizionali, come il confirmation bias, in cui il paziente tende inconsapevolmente a confermare le ipotesi del medico. Altre volte derivano dalla mancanza di dati relativi a specifici sottogruppi di popolazione: ad esempio, si pensi a un modello di IA diagnostica addestrata quasi esclusivamente su dati relativi a soggetti caucasici, il quale ha dimostrato prestazioni significativamente inferiori quando applicata a pazienti con tonalità di pelle diverse⁵³.

L'interoperabilità rappresenta un'ulteriore sfida nello scambio di informazioni, ad esempio tra fornitori di servizi sanitari e unità di ricerca. Diverse istituzioni impiegano strumenti differenti nei propri processi di gestione dei dati, rendendo l'integrazione delle informazioni provenienti da fonti multiple un'operazione particolarmente complessa e dispendiosa⁵⁴.

È proprio grazie alla sua capacità di superare simultaneamente questi molteplici ostacoli che la tecnologia dei dati sintetici si sta affermando come uno strumento ampiamente apprezzato nelle scienze biomediche⁵⁵. I dati sintetici offrono infatti il massimo livello di protezione dei dati, superando in modo significativo gli algoritmi di de-identificazione tradizionali⁵⁶. I generatori di dati sintetici identificano accuratamente le informazioni contenute nei dati clinici, li modellano stocasticamente e generano nuovi dati

⁴⁷ Per cosa si intende per RWD si veda Cowley, Andrea. "What is real world data?". CRC Australia. Clinical Research Corporation. Retrieved 8 May 2018.

⁴⁸ A tal proposito si veda E. GETZEN, L. UNGAR, D. MOWERY, X. JIANG, Q. LONG, *Mining for equitable health: Assessing the impact of missing data in electronic health records*, in *Journal of Biomedical Informatics*, 139, 2023, art. 104269. doi: 10.1016/j.jbi.2022.104269. Si veda anche C.D. NEWGARD, R.J. LEWIS, *Missing Data: How to Best Account for What Is Not Known*, in *JAMA*, 314(9), 2015, 940-941. doi: 10.1001/jama.2015.10516

⁴⁹ A.L. KOMAROFF, *The variability and inaccuracy of medical data*, in *Proceedings of the IEEE*, 67(9), 1979, 1196-1207.

⁵⁰ S.I. GOLDBERG, A. NIEMIERKO, A. TURCHIN, *Analysis of data errors in clinical research databases*, in *AMIA Annual Symposium Proceedings*, 2008, 242-246.

⁵¹ S.I. GOLDBERG, A. NIEMIERKO, A. TURCHIN, *op.cit.*, 242-246.

⁵² A. ALTHUBAITI, *Information bias in health research: definition, pitfalls, and adjustment methods*, in *Journal of Multidisciplinary Healthcare*, 2016, 211-217.

⁵³ S. HARRIS, *Racial bias exists in photo-based medical diagnosis despite AI help*, 2024.

⁵⁴ C. PENG, P. GOSWAMI, G. BAI, *A literature review of current technologies on health data integration for patient-centered health management*, in *Health Informatics Journal*, 26(3), 2020, 1926-1951. doi:10.1177/1460458219892387

⁵⁵ A tal proposito si veda B. VAN BREUGEL, T. LIU, D. OGLIC et al., *Synthetic data in biomedicine via generative artificial intelligence*, in *Nature Reviews Bioengineering*, 2, 2024, 991-1004. <https://doi.org/10.1038/s44222-024-00245-7>. Si veda anche A. GONZALES, G. GURUSWAMY, S.R. SMITH, *Synthetic data in health care: A narrative review*, in *PLOS Digital Health*, 2(1), 2023, art. e0000082. doi: 10.1371/journal.pdig.0000082.

⁵⁶ J. RUMBOLD, B. PIERSCONEK, *Contextual Anonymization for Secondary Use of Big Data in Biomedical Research: Proposal for an Anonymization Matrix*, in *JMIR Medical Informatics*, 6(4), 2018, art. e47. doi: 10.2196/medinform.7096.

sintetici a partire dai modelli così ottenuti⁵⁷. I generatori contemporanei basati su IA generativa sono in grado di produrre dati sintetici iper-realistici che conservano l'utilità informativa dei dati clinici. Tale efficacia è stata dimostrata in contesti clinici che spaziano dalle neoplasie ematologiche⁵⁸ all'epatologia⁵⁹, dall'oncologia⁶⁰ alle malattie cardiovascolari⁶², fino ad altri ambiti specialistici.

Oltre a ridurre drasticamente i tempi di accesso ai dati grazie a un'alternativa rispettosa della privacy e al contempo informativamente valida rispetto ai dati clinici, i dati sintetici contribuiscono alla disponibilità dell'evidenza e all'interoperabilità anche affrontando le sfide logistiche connesse all'accesso ai dati. Van Breugel ha dimostrato più volte come i dati sintetici siano in grado di superare i problemi legati ai dati mancanti e ai *bias*⁶³. Inoltre, i dati sintetici risultano immediatamente curati, formattati e coerenti. Strumenti come la piattaforma di dati sintetici sviluppata da Aindo⁶⁴ offrono interoperabilità diretta con i principali strumenti di gestione delle banche dati, semplificando il processo di generazione. In tal modo, la tecnologia dei dati sintetici rende l'informazione accessibile su larga scala e facilita l'interoperabilità tra i diversi attori della catena del valore dei dati sanitari.

Nel paragrafo successivo verranno esaminate più in dettaglio le opportunità offerte dai dati sintetici in termini di disponibilità dell'evidenza e interoperabilità, con riferimento a un ambito applicativo specifico: la valutazione comparativa.

3.2. Simulazione e confronto: il ruolo dei dati sintetici nella valutazione comparativa

I dati rappresentano l'infrastruttura fondamentale della valutazione comparativa e il potenziale dei dati sintetici in questo ambito non può essere sottovalutato. I dati sintetici possono essere impiegati nella valutazione di dispositivi medici e di altre tecnologie sanitarie. In questo contesto, la possibilità di simulare scenari tramite la generazione di dati sintetici assume un'importanza centrale. Durante il processo di generazione, infatti, è possibile modellare eventi specifici attraverso il riequilibrio delle coorti di

⁵⁷ J. JORDON, L. SZPRUCH, F. HOUSIAU et al., *Synthetic Data—what, why and how?*, in *arXiv preprint*, arXiv:2205.03257, 6 maggio 2022.

⁵⁸ A tal proposito si veda S. D'AMICO, D. DALL'OLIO, C. SALA et al., *Synthetic Data Generation by Artificial Intelligence to Accelerate Research and Precision Medicine in Hematology*, in *JCO Clinical Cancer Informatics*, 7, 2023, art. e2300021. doi: 10.1200/CCI.23.00021. Si veda anche S. D'AMICO, E. SAUTA, M. BERSANELLI et al., *Synthetic data generation by artificial intelligence to accelerate translational research and precision medicine in hematological malignancies*, in *Blood*, 140(Suppl. 1), 2022, 9744-9746.

⁵⁹ J.C. AHN, Y.K. NOH, M. HU et al., *AI-driven synthetic data generation for accelerating hepatology research: A study of the United Network for Organ Sharing (UNOS) database*, in *Hepatology*, 2025 (in stampa). doi: 10.1097/HEP.0000000000001299.

⁶⁰ R. KHALAF, W. DAVALAN, A. MOHAMMAD, R. DIAZ, *Synthetic data reliably reproduces brain tumor primary research data*, in *Canadian Journal of Neurological Sciences*, 51(s1), 2024, S41. doi:10.1017/cjn.2024.196.

⁶¹ R. LUN, D. SIEGAL, T. RAMSAY et al., *Synthetic data in cancer and cerebrovascular disease research: A novel approach to big data*, in *PLOS One*, 19(2), 2024, art. e0295921. doi: 10.1371/journal.pone.0295921.

⁶² Z. AZIZI, S. LINDNER, Y. SHIBA et al., *A comparison of synthetic data generation and federated analysis for enabling international evaluations of cardiovascular health*, in *Scientific Reports*, 13, 2023, art. 11540. <https://doi.org/10.1038/s41598-023-38457-3>.

⁶³ A tal proposito si veda B. VAN BREUGEL, LIU, OGLIC et al., *Synthetic data in biomedicine via generative artificial intelligence*, cit. Si veda anche B. VAN BREUGEL, M. VAN DER SCHAAR, *Beyond privacy: Navigating the opportunities and challenges of synthetic data*, in *arXiv preprint*, arXiv:2304.03722, 7 aprile 2023.

⁶⁴ <https://www.aindo.com/> (ultima consultazione 21/06/2025).



popolazione⁶⁵. Questo consente di verificare la validità e la robustezza delle nuove tecnologie anche in condizioni rare o impreviste.

Si consideri nuovamente l'esempio dell'IA diagnostica in dermatologia, addestrata prevalentemente su dati relativi a pazienti caucasici. Attraverso il sovracampionamento di dati sintetici con tonalità di pelle differenti, i valutatori possono testare l'affidabilità dello strumento su coorti più eterogenee. Nella parte restante di questa sezione, si analizzerà più nel dettaglio un esempio concreto del potenziale dei dati sintetici nella valutazione comparativa: le sperimentazioni cliniche.

Studi recenti indicano che i dati sintetici possono essere utilizzati in modo affidabile per sostituire, parzialmente o integralmente, i bracci di controllo nelle sperimentazioni cliniche⁶⁶. L'impatto di tale possibilità è difficilmente sopravvalutabile. Il reclutamento dei pazienti rappresenta oggi la voce di costo più rilevante nella ricerca clinica, assorbendo circa il 32% del budget totale⁶⁷, con una media di 6.500 dollari per paziente negli Stati Uniti (dati 2016⁶⁸). Oltre ai costi, il reclutamento è un processo lungo, che richiede in media 18 mesi⁶⁹.

L'utilizzo di bracci di controllo esterni, in cui il gruppo di controllo è sostituito da dati preesistenti, è divenuto una prassi consolidata per ridurre l'impiego di risorse. Tale approccio è riconosciuto e approvato sia dalla Food and Drug Administration (FDA⁷⁰) statunitense che dall'Agenzia Europea per i Medicinali (EMA)⁷¹. Tra il 2016 e il 2021⁷² il 17% dei farmaci oncologici approvati dall'EMA⁷³ ha fatto ricorso a bracci esterni; nel 63% dei casi, tali sperimentazioni sono state ritenute dall'EMA idonee a fornire prove a supporto dell'efficacia clinica.

Tuttavia, i bracci di controllo esterni presentano limiti significativi⁷⁴. Il riutilizzo reiterato di un numero ristretto di dataset standardizzati può renderli obsoleti e introdurre bias sistematici⁷⁵. Inoltre, i RWD

⁶⁵ B. VAN BREUGEL, VAN DER SCHAAR, *Beyond privacy: Navigating the opportunities and challenges of synthetic data*, cit.

⁶⁶ A tal proposito si veda S. EL KABABJI, N. MITSAKAKIS, X. FANG et al., *Evaluating the Utility and Privacy of Synthetic Breast Cancer Clinical Trial Data Sets*, in *JCO Clinical Cancer Informatics*, 7, 2023, art. e2300116. doi: 10.1200/CCI.23.00116. Si veda anche J.N. ECKARDT, W. HAHN, C. RÖLLIG et al., *Mimicking clinical trials with synthetic acute myeloid leukemia patients using generative artificial intelligence*, in *NPJ Digital Medicine*, 7, 2024, art. 76. <https://doi.org/10.1038/s41746-024-01076-x>. Si veda anche S. ELVATUN, D. KNOORS, S. BRANT et al., *Synthetic data as external control arms in scarce single-arm clinical trials*, in *PLOS Digital Health*, 4(1), 2025, art. e0000581. <https://doi.org/10.1371/journal.pdig.0000581>.

⁶⁷ DELOITTE UNIVERSITY EMEA CVBA, *Deloitte Insights, Intelligent Clinical Trials*, 2020.

⁶⁸ T.J. MOORE, H. ZHANG, G. ANDERSON, G.C. ALEXANDER, *Estimated Costs of Pivotal Trials for Novel Therapeutic Agents Approved by the US Food and Drug Administration, 2015–2016*, in *JAMA Internal Medicine*, 178(11), 2018, 1451–1457. <https://doi.org/10.1371/journal.pdig.0000581>

⁶⁹ M. BRØGGER-MIKKELSEN, J.R. ZIBERT, A.D. ANDERSEN et al., *Changes in key recruitment performance metrics from 2008–2019 in industry-sponsored phase III clinical trials registered at ClinicalTrials.gov*, in *PLOS One*, 17(7), 2022, art. e0271819. doi: 10.1371/journal.pone.0271819.

⁷⁰ US FOOD AND DRUG ADMINISTRATION, *Framework for FDA's Real-World Evidence Program*, 2018, 20.

⁷¹ X. WANG, F. DORMONT, C. LORENZATO et al., *Current perspectives for external control arms in oncology clinical trials: Analysis of EMA approvals 2016–2021*, in *Journal of Cancer Policy*, 35, 2023, art. 100403. doi: 10.1016/j.jcpo.2023.100403.

⁷² X. WANG, F. DORMONT, C. LORENZATO et al., op. cit.

⁷³ X. WANG, F. DORMONT, C. LORENZATO et al., op. cit.

⁷⁴ R. BANERJEE, S. MIDHA, A.H. KELKAR et al., *Synthetic control arms in studies of multiple myeloma and diffuse large B-cell lymphoma*, in *British Journal of Haematology*, 196(5), 2022, 1274–1277. doi: 10.1111/bjh.17945.

⁷⁵ J.P. LYMAN, A. DOUCETTE, B. ZHENG-LIN et al., *Feasibility and utility of synthetic control arms derived from real-world data to support clinical development*, [incomplete – manca sede di pubblicazione].

utilizzati possono contenere variabili non misurate, come comorbidità, che escluderebbero quei pazienti da una sperimentazione effettiva, conducendo così a risultati potenzialmente fuorvianti⁷⁶.

I dati sintetici offrono una risposta diretta a queste criticità, ampliando le possibilità di realizzare analisi comparative più efficaci. La possibilità di generare dati sintetici aggiornati consente di produrre evidenze accurate e prive di bias, rispondenti alle esigenze specifiche dei ricercatori⁷⁷.

3.3. Verso una nuova grammatica dell'HTA basata sull'evidenza

Alla luce del potenziale delineato, non sorprende che la tecnologia dei dati sintetici abbia suscitato un crescente interesse tra gli attori del settore dei dati sanitari, i legislatori e i decisori politici. In collaborazione con le altre Autorità per la Protezione dei Dati (DPA) dei Paesi del G7, l'Information Commissioner's Office (ICO) del Regno Unito ha pubblicato un caso d'uso sull'impiego dei dati sintetici nel processo decisionale in ambito di sanità pubblica⁷⁸.

Il caso esaminato riguarda un governo regionale incaricato di distribuire scorte mediche tra i vari erogatori di servizi sanitari del territorio. Per realizzare un'allocazione ottimale delle risorse, l'amministrazione ha bisogno di dati accurati sulla domanda proveniente da ciascun fornitore. Ciò consente di comprendere quali quantità di medicinali siano necessarie in ciascuna area. Tuttavia, tali informazioni risultano altamente sensibili, in quanto attinenti direttamente ai bisogni terapeutici dei pazienti. I dati sintetici vengono quindi utilizzati per rendere accessibili al governo regionale le informazioni sulla domanda, senza tuttavia condividere dati sensibili.

Questo caso d'uso elaborato dall'ICO dimostra concretamente quanto i dati sintetici contribuiscano alla disponibilità dell'evidenza, come già illustrato nei paragrafi precedenti. Non solo i dati diventano rapidamente accessibili, ma la loro disponibilità sicura, accurata ed efficiente consente anche l'aggregazione di fonti altrimenti eterogenee. Nell'esempio citato, l'aggregazione dei dati sintetici permette al governo regionale di prendere decisioni che coinvolgono simultaneamente più strutture sanitarie. Si tratta di un chiaro esempio in cui l'intero, ovvero il dataset sintetico aggregato, è superiore alla somma delle sue parti. Questa capacità di aggregazione offerta dai dati sintetici ha il potenziale per ridefinire la stessa grammatica dell'HTA basata sull'evidenza. Un caso emblematico di questa nuova prospettiva riguarda le malattie rare. Per loro natura, esse sono caratterizzate da una scarsità strutturale di dati. È dunque particolarmente complesso raccogliere quantità sufficienti di dati clinici utili alla progettazione o alla valutazione di una tecnologia sanitaria. I dati sintetici offrono l'opportunità di raccogliere e aggregare informazioni da una pluralità di istituzioni, superando così la barriera della scarsità e aprendo nuovi scenari per la valutazione scientifica e regolatoria. Quest'ultimo aspetto sarà approfondito dettagliatamente da una prospettiva clinica nel primo capitolo della seconda parte del lavoro, la quale è dedicata all'analisi delle implicazioni cliniche, etiche, economiche e filosofiche dell'impiego dei dati sintetici nella sperimentazione clinica.

⁷⁶ R. BANERJEE et al., *op. cit.*

⁷⁷ S. EL KABABJI et al., *op. cit.*; S. ELVATUN et al., *op. cit.*

⁷⁸ <https://ico.org.uk/.../g7-dpas-emerging-technologies-working-group-use-case-study-on-privacy-enhancing-technologies/> (ultima consultazione 21/06/2025).

