



TINA SALERNO

Implementazione di un modello multidimensionale per il bilanciamento dei diritti nelle investigazioni AI-based ad alto impatto

Il contributo propone PEGA 2, modello operativo e multidimensionale per valutare la proporzionalità nelle investigazioni AI-based. Aggiornamento metodologico del framework PEGA, tratto dalla Raccomandazione della Commissione d'inchiesta del Parlamento europeo sull'uso di spyware (2022-2023), ne conserva l'impianto convenzionale integrandolo con parametri tecnici, socio-economici e giuridici. Elemento distintivo è l'Indice di Intrusività Costituzionalmente Rilevante (IICR), che collega gravità della minaccia (M), affidabilità tecnica e probatoria del sistema (A) e impatto socio-economico della misura (S). Le variabili poggiano sulla giurisprudenza della Corte EDU su accesso ai dati, sorveglianza segreta e non-disclosure e sui profili di opacità algoritmica, rischio cibernetico e profilazione predittiva. Il modello fornisce criteri operativi per determinarne i valori e affronta l'interdipendenza tra stima algoritmica della minaccia e affidabilità del sistema, distinguendo investigazioni *ex post* e predittive.

*Intelligenza artificiale investigativa – Proporzionalità – Sorveglianza digitale – Diritti fondamentali
Cybersecurity*

Implementation of a multidimensional model for balancing rights in high-impact AI-based investigations

This article proposes PEGA 2, an operational, multidimensional model for assessing proportionality in AI-based investigations. A methodological update of the PEGA framework, drawn from the Recommendation closing the European Parliament's spyware inquiry (2022-2023), it keeps its conventional basis, adding technical, socio-economic and legal parameters. Its core is the Constitutionally Relevant Intrusiveness Index (IICR), relating threat gravity (M), technical and evidentiary reliability of the system (A) and the measure's socio-economic impact (S). The variables rest on ECtHR case law on data access, secret surveillance and non-disclosure, and on algorithmic opacity, cyber risk and predictive profiling. It sets operational criteria for assigning their values and addresses the interdependence between algorithmic threat estimation and system reliability, distinguishing *ex post* from predictive investigations.

Investigative artificial intelligence – Proportionality – Digital surveillance – Fundamental rights – Cybersecurity

SOMMARIO: 1. Introduzione. – 2. Il framework PEGA: origini, contenuto e limiti. – 3. Il modello PEGA 2: struttura, variabili e Indice di Intrusività Costituzionalmente Rilevante. – 3.1. Criteri operativi per la misurazione delle variabili. – 3.2. Parametri predeterminati e margini di valutazione. – 3.3. L'interdipendenza tra *M* e *A*: regime *ex post* e regime predittivo. – 4. Accesso ai dati personali e limiti convenzionali. – 5. Trasparenza e controllo indipendente. – 6. Giurisprudenza della Corte EDU e non-disclosure di dati personali. – 7. Opacità algoritmica e crittografia nelle investigazioni AI-based. – 8. Cybersecurity e rischio investigativo. – 9. Profilazione predittiva e intrusività anticipatoria. – 10. Affidabilità probatoria come sotto-indicatore autonomo. – 11. Dal framework PEGA a PEGA 2: continuità e aggiornamento. – 12. Conclusioni.

1. Introduzione

Il ricorso a sistemi informatici basati sull'intelligenza artificiale nell'ambito delle attività investigative¹ ha conosciuto, negli ultimi anni, un'espansione significativa, favorita dalla crescente disponibilità di dati, dall'evoluzione delle tecniche di analisi algoritmica e dall'intensificarsi delle forme di cooperazione giudiziaria e di polizia a livello transnazionale². Piattaforme di analisi predittiva, sistemi di profilazione automatizzata, software di correlazione massiva dei dati e strumenti di supporto decisionale sono progressivamente impiegati non solo come ausili investigativi, ma come dispositivi capaci di incidere in modo strutturale sulle modalità di accertamento dei fatti e di selezione dei soggetti di interesse³.

Tale evoluzione si colloca all'interno di un più ampio processo di digitalizzazione della sorveglianza, nel quale la raccolta e il trattamento dei

dati personali avvengono sempre più frequentemente in forme pervasive, opache e difficilmente percepibili dall'interessato⁴. L'impiego di sistemi informatici di intelligenza artificiale introduce un mutamento qualitativo rispetto alle tecniche investigative tradizionali: consente non solo di ampliare la scala dell'osservazione, ma anche di anticipare l'intervento repressivo attraverso meccanismi di inferenza e previsione del comportamento, producendo un'intrusività che non si manifesta più attraverso un atto puntuale e identificabile, bensì attraverso la combinazione di trattamenti automatizzati, correlazioni invisibili e decisioni opache, distribuite lungo catene transnazionali di cooperazione⁵.

Le categorie giuridiche classiche elaborate per valutare la legittimità delle misure di sorveglianza – fondate su concetti quali la materialità dell'ingerenza, la delimitazione temporale e spaziale

1. Parte della presente ricerca è stata svolta durante il periodo in cui l'autrice è stata ospitata dalla Direzione Nazionale Antimafia e Antiterrorismo per finalità di ricerca e formazione, nell'ambito del proprio percorso dottorale e in conformità al relativo accordo di ospitalità. Durante la permanenza presso l'Istituzione ospitante, l'autrice ha beneficiato del supporto e della supervisione del Sostituto Procuratore Nazionale Antonio Ardituro. Le opinioni espresse sono esclusivamente riconducibili all'autrice e non riflettono necessariamente quelle della Direzione Nazionale Antimafia e Antiterrorismo.

2. CAMALDO 2025; VASTA 2025.

3. TADDEI ELMI–MARCHIAFAVA 2024.

4. CARAMASCHI 2025; SIMONCINI 2019; HILDEBRANDT 2015.

5. POLLICINO 2025; KOOPS–LEENES 2014.

del controllo e la riconoscibilità dell'atto invasivo – appaiono sempre meno adeguate a cogliere l'impatto effettivo di questo tipo di tecnologie informatiche sui diritti fondamentali. Ne deriva l'esigenza di ripensare i criteri attraverso i quali valutare la compatibilità delle pratiche investigative basate sull'intelligenza artificiale con i limiti imposti dalla Convenzione europea dei diritti dell'uomo e dalla normativa europea in materia di protezione dei dati personali, con particolare riferimento al principio di proporzionalità quale parametro cardine di legittimazione dell'ingerenza nei diritti individuali.

Un primo tentativo di risposta a questa esigenza è stato offerto dal *framework PEGA*: uno schema qualitativo di valutazione della proporzionalità delle misure investigative ad alto impatto, elaborato a partire dalla Raccomandazione adottata dal Parlamento europeo il 15 giugno 2023 in esito ai lavori della Commissione parlamentare di inchiesta sull'uso di Pegasus e di software di sorveglianza equivalenti (Commissione PEGA, 2022-2023)⁶. Quel framework ha il merito di aver tradotto in criteri operativi – autorizzazione giudiziaria *ex ante*, proporzionalità e necessità, supervisione indipendente *ex post*, notifica all'interessato, cancellazione dei dati non pertinenti – le indicazioni della Corte europea dei diritti dell'uomo in materia di sorveglianza. Tuttavia, nato in risposta a un contesto specifico, quello degli spyware commerciali, il framework PEGA presenta limiti strutturali che ne circoscrivono l'applicabilità al più ampio scenario delle investigazioni basate su sistemi informatici di intelligenza artificiale: è essenzialmente qualitativo e formale; non considera la complessità tecnica dei sistemi AI, i rischi di bias algoritmico, la resilienza

cibernetica e l'affidabilità probatoria dei risultati; non tiene conto dell'impatto socio-economico e istituzionale delle misure investigative; non prevede strumenti operativi per gestire l'interoperabilità transnazionale tra sistemi di ordinamenti diversi.

Il presente contributo muove da questi limiti per proporre un aggiornamento metodologico del framework, qui denominato *PEGA 2*, che ne conserva l'impianto convenzionale integrandolo con i parametri tecnici, socio-economici e giuridici che quel modello non contemplava. L'elemento distintivo della proposta è l'*Indice di Intrusività Costituzionalmente Rilevante (IICR)*, uno strumento analitico – da intendersi non come algoritmo decisionale definitivo, bensì come quadro strutturato di valutazione – che traduce il principio di proporzionalità nella relazione tra gravità della minaccia criminale (*M*), affidabilità tecnica e probatoria del sistema informatico impiegato (*A*) e impatto socio-economico della misura (*S*). La costruzione di ciascuna variabile trae fondamento dall'analisi della giurisprudenza della Corte europea dei diritti dell'uomo e dalle specificità tecniche dei sistemi AI investigativi, secondo una logica per cui ogni sezione del contributo non è accessoria rispetto al modello, ma ne costituisce il fondamento argomentativo⁷.

Il contributo si colloca pienamente nell'ambito dell'informatica giuridica, intesa come spazio di intersezione tra analisi giuridica e comprensione delle infrastrutture tecnologiche⁸. L'ipotesi di fondo è che l'impiego di sistemi informatici basati sull'intelligenza artificiale nelle investigazioni non possa essere valutato unicamente alla luce dei criteri tradizionali di necessità e adeguatezza della misura, ma richieda l'introduzione del concetto

6. Parlamento europeo, *Raccomandazione del Parlamento europeo del 15 giugno 2023 al Consiglio e alla Commissione a seguito dell'indagine sull'uso di Pegasus e di spyware di sorveglianza equivalenti* (2023/2500(RSP)), B9-0260/2023, P9_TA(2023)0244, §§ 30-32; Relazione A9-0189/2023.

7. La proposta qui presentata matura a partire da un percorso di ricerca multidisciplinare che ha incluso un periodo di osservazione sul campo di sei mesi presso il Justice Digital Center di Wrocław (Polonia). Tale esperienza ha consentito di confrontarsi con prassi operative concrete e di costruire un dataset esplorativo di cinquanta scenari investigativi (reali e simulati) che ha orientato la definizione delle variabili e delle soglie del modello. I risultati vanno intesi come contributo preliminare e base per ulteriori verifiche empiriche.

8. FAMELI 2025; FALLETTA-MARSANO 2024; FINOCCHIARO 2024; MARTIRE 2024; TADDEI ELMI-CONTALDO 2024; BRIGHI 2022; PALMIRANI-SAPIENZA 2022; SANTOSUOSSO-SARTOR 2022; SARTOR 2022-A; SARTOR 2022-B, pp. 1-30; ZICCARDI 2022; MANTELERO-ESPOSITO 2021; MANES 2020; SIMONCINI 2019; ZICCARDI-PERRI 2019; BRIGHI-PALMIRANI-SÁNCHEZ JORDÁN 2018; CEPEJ 2018; HILDEBRANDT 2015; KOOPS-LEENES 2014; PAGALLO 2013; COGNETTI 2011.

di *proporzionalità operativa*: un criterio dinamico di bilanciamento che tiene conto non solo dell'obiettivo investigativo perseguito, ma anche delle modalità concrete di funzionamento dei sistemi utilizzati, del grado di automazione decisionale, dell'opacità algoritmica, del rischio cibernetico e dell'affidabilità probatoria dei risultati prodotti⁹.

Il contributo è strutturato come segue. Il paragrafo 2 illustra il framework PEGA, ne descrive i criteri qualitativi di proporzionalità e ne identifica i limiti strutturali che motivano il presente aggiornamento. Il paragrafo 3 presenta la struttura di PEGA 2 e l'IICR, definendo le variabili M, A e S, la logica della formula, i criteri operativi per la misurazione delle variabili (par. 3.1), la natura – predeterminata o valutativa – dei parametri impiegati (par. 3.2) e il trattamento dell'interdipendenza tra M e A (par. 3.3). I paragrafi 4, 5 e 6 analizzano la giurisprudenza della Corte europea dei diritti dell'uomo in materia di accesso ai dati, trasparenza e non-disclosure, che costituisce il fondamento convenzionale delle variabili del modello. I paragrafi 7, 8 e 9 esaminano i profili tecnico-giuridici – opacità algoritmica, rischio cibernetico e profilazione predittiva – che determinano concretamente il valore della variabile A; il paragrafo 9 sviluppa inoltre, per il regime predittivo, le regole operative formulate al paragrafo 3.3. Il paragrafo 10 affronta l'affidabilità probatoria come sotto-indicatore autonomo. Il paragrafo 11 riepiloga il confronto tra PEGA e PEGA 2 e illustra la struttura del dataset esplorativo attraverso alcuni scenari-tipo. Il paragrafo 12 trae le conclusioni.

2. Il framework PEGA: origini, contenuto e limiti

Il framework PEGA trae origine dai lavori della Commissione parlamentare di inchiesta del Parlamento europeo sull'uso di Pegasus e di software di sorveglianza equivalenti (Commissione PEGA), istituita nel marzo 2022 e conclusasi con l'adozione del rapporto finale nel maggio 2023 e della raccomandazione plenaria il 15 giugno 2023¹⁰. Nel corso di quattordici mesi di audizioni, missioni

conoscitive e studi commissionati, la Commissione ha analizzato l'impiego di spyware commerciali – in primo luogo Pegasus del gruppo NSO – da parte di governi di Stati membri dell'Unione europea nei confronti di giornalisti, avvocati, politici e attivisti, accertando violazioni sistematiche dei diritti fondamentali e dell'ordinamento europeo, in particolare in Ungheria e Polonia.

Sul piano normativo, la Raccomandazione B9-0260/2023 ha elaborato cinque criteri qualitativi per la valutazione della legittimità delle misure di sorveglianza mediante spyware: (i) autorizzazione giudiziaria *ex ante* efficace, vincolante e significativa (§ 32.a); (ii) proporzionalità e necessità rispetto alla gravità del reato perseguito, limitata a reati gravi e minacce reali alla sicurezza nazionale (§ 32.c); (iii) supervisione indipendente *ex post* con poteri effettivi di intervento (§ 32.j); (iv) obbligo di notifica all'interessato al termine della sorveglianza (§ 32.h); (v) cancellazione di tutti i dati non pertinenti durante e dopo l'indagine (§ 32.p). Questi criteri – fondati sulla giurisprudenza della Corte europea dei diritti dell'uomo (Corte EDU) e sulle garanzie previste dal diritto dell'Unione europea in materia di protezione dei dati personali – costituiscono il nucleo del framework PEGA.

Il framework PEGA rappresenta un contributo metodologico significativo: ha offerto una prima traduzione operativa dei principi convenzionali in criteri applicabili alle misure investigative ad alto impatto, colmando un vuoto nell'ambito della valutazione della proporzionalità. Nondimeno, il suo perimetro di applicazione è strettamente connesso al contesto che lo ha generato – quello degli spyware commerciali – e i suoi criteri riflettono i limiti di quel contesto specifico.

In primo luogo, i criteri elaborati dalla Commissione sono di natura esclusivamente qualitativa e formale: verificano la legalità procedurale della misura, ma non forniscono strumenti per misurare l'intensità effettiva dell'ingerenza nei diritti fondamentali né per ponderarla rispetto alla gravità della minaccia investigativa. In secondo luogo, il framework non considera la complessità tecnica dei

9. FALLETTA-MARSANO 2024; FINOCCHIARO 2024; MARTIRE 2024; TADDEI ELMI-MARCHIAFAVA 2024; SARTOR 2022-A.

10. Raccomandazione del Parlamento europeo del 15 giugno 2023, P9_TA(2023)0244 (B9-0260/2023); Parlamento europeo, *Relazione sull'indagine sull'uso di Pegasus e di software di sorveglianza equivalenti* (relatrice Sophie in 't Veld), A9-0189/2023, approvata dalla Commissione d'inchiesta PEGA l'8 maggio 2023.

sistemi informatici di intelligenza artificiale impiegati nelle indagini – il rischio di bias algoritmico, la probabilità di falsi positivi, la resilienza cibernetica, l'affidabilità probatoria dei risultati – elementi che incidono direttamente sull'intensità reale dell'ingerenza. In terzo luogo, non tiene conto dell'impatto socio-economico e istituzionale delle misure investigative, né delle sfide poste dall'interoperabilità transnazionale tra sistemi di ordinamenti diversi. È a partire da questi limiti strutturali – peraltro comprensibili data la genesi del framework – che il presente contributo propone PEGA 2.

3. Il modello PEGA 2: struttura, variabili e Indice di Intrusività Costituzionalmente Rilevante

PEGA 2 è una proposta di aggiornamento metodologico del framework PEGA, concepita per estenderne l'applicabilità al più ampio scenario delle investigazioni basate su sistemi informatici di intelligenza artificiale. Il modello conserva l'impianto convenzionale del framework originario – fondato sulla giurisprudenza della Corte europea dei diritti dell'uomo e sui principi europei di protezione dei dati – integrandolo con tre dimensioni operative che il framework PEGA non contemplava.

La prima dimensione è la *robustezza tecnica e affidabilità probatoria* dei sistemi informatici impiegati: comprende la verifica di bias e falsi positivi, l'analisi dei margini di errore, la resilienza a cyber-attacchi e la documentazione dell'architettura e dei dati di addestramento. Quanto più un sistema è suscettibile di errori o manipolazioni, tanto maggiore è l'interferenza reale sui diritti individuali, indipendentemente dalla legittimità formale della misura. La seconda dimensione è l'*impatto socio-economico e istituzionale* delle misure investigative, valutato sia nei suoi effetti diretti (perdite economiche, danni ai mercati, impatto sulle finanze pubbliche) sia indiretti (fiducia nelle istituzioni, coesione sociale). La terza dimensione riguarda la *validazione giuridica, trasparenza e interoperabilità internazionale*: integra la conformità a normative quali il GDPR, la Direttiva LED (Law Enforcement Directive) e il Regolamento sull'intelligenza artificiale, e richiede supervisione umana delle decisioni algoritmiche, tracciabilità dei processi e

gestione dei rischi di disallineamento normativo tra ordinamenti diversi.

Il cuore operativo di PEGA 2 è l'*Indice di Intrusività Costituzionalmente Rilevante* (ICR), definito dalla formula:

$$I_{ICR} = M / [A \times (1 - S)]$$

dove *M* rappresenta la gravità della minaccia criminale (scala 0-1), *A* indica l'affidabilità tecnica e probatoria del sistema informatico impiegato – comprensiva di assenza di bias, resilienza cibernetica e trasparenza algoritmica (scala 0-1) – e *S* esprime l'impatto socio-economico e istituzionale della misura (scala 0-1). L'indice cresce al crescere della gravità della minaccia e dell'impatto sociale, e decresce al crescere dell'affidabilità del sistema. Quando *A* tende a zero – sistema opaco, inaffidabile o vulnerabile – l'indice supera qualunque soglia di ammissibilità costituzionale indipendentemente dall'obiettivo investigativo: la formula esprime così, in termini analitici, il principio per cui l'inaffidabilità tecnica non è un aspetto neutro, ma un fattore che aggrava strutturalmente l'intrusività della misura.

L'ICR non è uno strumento di calcolo automatico della legittimità di una misura investigativa, né un algoritmo decisionale: è un quadro strutturato di valutazione che impone di esplicitare e ponderare variabili che nei modelli tradizionali rimangono implicite o ignorate. La sua utilità consiste nell'obbligare chi adotta o controlla una misura investigativa AI-based a documentare il valore di *A* – quanto è affidabile il sistema usato? – e a considerare *S* – quali ricadute sociali ed economiche produce? – accanto alla più tradizionale valutazione della gravità del reato (*M*).

La costruzione delle variabili e la logica della formula trovano fondamento nell'analisi giurisprudenziale e tecnico-normativa sviluppata nei successivi paragrafi 4-10: ogni sezione identifica come le indicazioni della Corte europea dei diritti dell'uomo e le specificità tecniche dei sistemi AI investigativi determinino concretamente il valore di *M*, *A* e *S*. L'obiettivo non è produrre un sistema chiuso, ma offrire un'architettura concettuale che orienti la valutazione della proporzionalità in un contesto tecnologico per il quale le categorie giuridiche classiche risultano insufficienti¹¹.

11. SARTOR 2022-B; MANTELERO-ESPOSITO 2021; HILDEBRANDT 2015.

3.1. Criteri operativi per la misurazione delle variabili

Perché l'IICR non si risolva in un'enunciazione di principio – capace di indicare che cosa pesare, ma non come pesarlo – occorre che i valori di *M*, *A* e *S* non siano assegnati in modo intuitivo, bensì determinati attraverso criteri espliciti, documentabili e controllabili. Questo paragrafo propone, per ciascuna variabile, una griglia operativa di prima approssimazione: criteri suscettibili di affinamento nelle verifiche empiriche future, ma sufficienti a rendere applicabile il modello e a vincolare la motivazione di chi lo impiega.

La variabile *M* è ancorata, in via prioritaria, a parametri normativi predeterminati. La griglia adottata articola la scala 0-1 in quattro fasce: nella fascia inferiore (0-0,25) si collocano i reati di minore gravità, estranei ai cataloghi europei della criminalità grave; nella seconda fascia (0,26-0,50) i reati di media gravità; nella terza (0,51-0,75) la criminalità grave riconducibile ai cataloghi normativi europei – in particolare l'art. 2, par. 2, della decisione quadro 2002/584/GAI sul mandato d'arresto europeo e l'allegato II della direttiva (UE) 2016/681 – e alla nozione di criminalità "grave" elaborata dalla Corte di giustizia in materia di conservazione dei dati; nella fascia superiore (0,76-1) la criminalità organizzata transnazionale, il terrorismo e le minacce attuali e concrete alla sicurezza nazionale¹². All'ancoraggio normativo si affiancano due correttivi valutativi: l'attualità e concretezza della minaccia, che può ridurre il valore entro la fascia di appartenenza, e la distinzione tra minaccia accertata *ex post* e minaccia stimata in via predittiva, che rinvia al regime speciale delineato al paragrafo 3.3.

La variabile *A* è costruita come media ponderata di quattro sotto-indicatori, ciascuno espresso su scala 0-1: (a) l'*accuratezza statistica documentata* del sistema, misurata attraverso precisione, recall e tasso di falsi positivi rilevati su dataset di validazione indipendenti da quelli di addestramento; (b) la *trasparenza e documentazione*, intesa come

effettiva disponibilità agli organi di controllo della documentazione tecnica richiesta dagli artt. 11-13 del Regolamento sull'intelligenza artificiale – architettura, dati di addestramento, logiche decisionali; (c) la *resilienza cibernetica*, attestata dall'esito di risk assessment, audit di vulnerabilità e conformità agli standard di sicurezza di riferimento (par. 8); (d) l'*affidabilità probatoria*, valutata secondo gli indicatori sviluppati al paragrafo 10 – tracciabilità del flusso dei dati, riproducibilità dei risultati, revisione indipendente¹³. Un elemento metodologico è decisivo e va sottolineato sin d'ora: *A* si misura *ex ante* e sul sistema, mediante validazione condotta su basi informative indipendenti dal caso concreto, e non *ex post* sull'esito della singola indagine. Questa scelta, apparentemente tecnica, è ciò che consente al modello di affrontare l'interdipendenza tra *M* e *A* esaminata al paragrafo 3.3.

La variabile *S* combina indicatori diretti – costi economici quantificabili, effetti su mercati e finanze pubbliche – e indiretti – fiducia nelle istituzioni, effetti dissuasivi sull'esercizio di libertà fondamentali, coesione sociale – operazionalizzati attraverso proxy osservabili: il numero di soggetti terzi i cui dati sono trattati dalla misura, l'estensione temporale del trattamento, la reversibilità degli effetti prodotti, l'ampiezza della platea potenzialmente profilata e l'incidenza su categorie che godono di tutele rafforzate, quali giornalisti, avvocati e difensori dei diritti umani. Anche per *S* la scala 0-1 è articolata in fasce, con obbligo di motivare l'assegnazione in relazione a ciascun proxy rilevante.

3.2. Parametri predeterminati e margini di valutazione

La natura dei parametri appena descritti non è omogenea, ed è opportuno esplicitarlo. Il modello distingue tre livelli. Al primo livello si collocano i parametri *predeterminati* in via eteronoma: i cataloghi normativi dei reati gravi, gli obblighi documentali fissati dal Regolamento sull'intelligenza artificiale, le metriche statistiche standard, gli standard tecnici di sicurezza. Rispetto a essi il

12. Sulla nozione di criminalità "grave" elaborata dalla Corte di giustizia in materia di conservazione dei dati, cfr. Corte di giustizia dell'Unione europea, Grande Sezione, 30 aprile 2024, causa C-178/22, *Procura della Repubblica presso il Tribunale di Bolzano*.

13. Regolamento (UE) 2024/1689 (AI Act), artt. 11-13, sugli obblighi di documentazione tecnica e di trasparenza per i sistemi ad alto rischio; sulla metodologia di valutazione d'impatto, MANTELERO-ESPOSITO 2021; DANEZIS-DOMINGO-FERRER-HANSEN et al. 2014.

valutatore accerta, non sceglie. Al secondo livello stanno i parametri *semi-strutturati*: le griglie a fasce descritte al paragrafo 3.1, i cui descrittori qualitativi devono essere applicati al caso concreto attraverso un'assegnazione motivata. Al terzo livello permane un *margin* di apprezzamento residuo, ineliminabile.

Questo margine non costituisce un difetto occulto del modello, né la natura aperta del sistema lo trasforma in arbitrio: esso è la trasposizione – resa esplicita e per ciò stesso controllabile – della discrezionalità che già caratterizza ogni giudizio di proporzionalità. L'IICR non elimina tale discrezionalità, la disciplina: impone la motivazione analitica di ciascun valore assegnato, espone l'assegnazione al controllo dell'autorità giudiziaria e degli organi di supervisione indipendente, e si presta – nelle sperimentazioni future – a verifiche di concordanza tra valutatori distinti sui medesimi scenari¹⁴. La natura aperta del sistema opera, inoltre, al livello delle griglie e non delle singole valutazioni: i parametri possono essere aggiornati in via generale e astratta al mutare del quadro normativo e tecnologico, ma nel singolo caso il valutatore è vincolato alla griglia vigente.

3.3. L'interdipendenza tra *M* e *A*: regime *ex post* e regime predittivo

Il modello deve confrontarsi con un'obiezione strutturale. Se, negli scenari predittivi, la gravità della minaccia deve essere stimata sulla base di inferenze algoritmiche, *M* non rappresenta un dato esperienziale derivante dall'osservazione, ma un output prodotto dallo stesso sistema la cui affidabilità è misurata da *A*: le due variabili non sarebbero allora indipendenti, e la formula rischierebbe di poggiare su una stima che il sistema produce di sé.

La risposta del modello si articola in una distinzione di regime. Nelle investigazioni *ex post* – che costituiscono la maggior parte degli scenari considerati – il reato è già commesso, o il sospetto si fonda su elementi raccolti con mezzi investigativi

tradizionali: *M* è un dato esperienziale, determinato in modo indipendente dal sistema informatico impiegato, e nessuna circolarità si produce. Nelle investigazioni *predittive*, la dipendenza di *M* da *A* è invece reale, ma – questo il punto – è unidirezionale e gerarchica, non circolare, a condizione che *A* sia misurata su una base informativa distinta. È precisamente la funzione della scelta metodologica enunciata al paragrafo 3.1: *A* è determinata *ex ante*, mediante validazione del sistema su dati indipendenti dal caso concreto – test set esterni, audit retrospettivi sulla correttezza di predizioni passate rispetto a riscontri accertati – sicché *M* dipende da *A*, ma *A* non dipende da *M*. La stima algoritmica della minaccia è condizionata dall'affidabilità del sistema; l'affidabilità del sistema non è condizionata dalla stima che esso produce nel caso di specie.

Da questa impostazione discendono tre regole operative per il regime predittivo. La prima è una condizione di ammissibilità: la formula si applica solo se *A* supera una soglia rafforzata di validazione ($A \geq A^*$); al di sotto di essa la misura predittiva è inammissibile a prescindere dal valore stimato di *M*, perché un sistema non validato non può fornire la base conoscitiva di un'ingerenza anticipatoria. La seconda è una regola di stima prudenziale: *M* entra nella formula con il valore inferiore dell'intervallo di stima prodotto dal sistema, così che l'incertezza operi a favore del soggetto inciso e non dell'ingerenza. La terza è una regola di corroborazione esterna: la stima algoritmica della minaccia deve trovare riscontro in elementi investigativi indipendenti dal sistema prima che la misura sia adottata o prorogata – trasposizione, sul piano del modello, dell'esigenza convenzionale che l'autorizzazione delle misure di sorveglianza si fondi su un sospetto ragionevole suscettibile di verifica¹⁵. L'interdipendenza tra *M* e *A* non è dunque rimossa: è riconosciuta e incorporata nel modello come vincolo di ammissibilità, secondo la logica sviluppata, sul piano applicativo, al paragrafo 9.

14. Sulla struttura del giudizio di proporzionalità e sul contenimento della discrezionalità mediante motivazione analitica, cfr. COGNETTI 2011; per i principi di qualità e controllo nell'impiego dell'intelligenza artificiale nei sistemi giudiziari, CEPEJ 2018.

15. Corte EDU, Grande Camera, *Roman Zakharov c. Russia*, ric. n. 47143/06, 4 dicembre 2015, sull'esigenza che l'autorizzazione delle misure di sorveglianza si fondi su un sospetto ragionevole suscettibile di verifica; nello stesso senso Corte EDU, *Szabó e Vissy c. Ungheria*, ric. n. 37138/14, 12 gennaio 2016.

4. Accesso ai dati personali e limiti convenzionali

Nella giurisprudenza della Corte europea dei diritti dell'uomo, l'accesso ai dati personali detenuti dalle autorità pubbliche è stabilmente qualificato come una componente essenziale del diritto al rispetto della vita privata ai sensi dell'art. 8 CEDU. Il diritto dell'individuo a conoscere l'esistenza, il contenuto e le modalità di trattamento dei dati che lo riguardano non si configura come una mera prerogativa procedurale, bensì come garanzia sostanziale contro forme occulte o sproporzionate di ingerenza statale.

A partire da *Gaskin c. Regno Unito*, la Corte ha riconosciuto che l'accesso a documenti e informazioni detenuti dallo Stato può costituire un interesse vitale protetto dalla Convenzione, strumentale alla comprensione della propria storia personale e alla tutela dell'identità individuale. Tale orientamento è stato consolidato da pronunce successive: *Rotaru c. Romania*, *Haralambie c. Romania*, *Jarnea c. Romania*, *Joanna Szulc c. Polonia* e *Yonchev c. Bulgaria* attestano come la mera conservazione di informazioni personali in archivi statali, unita all'impossibilità per l'interessato di accedervi o verificarne l'esattezza, costituisca di per sé un'ingerenza rilevante ai sensi della Convenzione¹⁶. Questi principi si estendono al settore di polizia, dove strumenti del Consiglio d'Europa – dalla Raccomandazione R (87)15 alla Convenzione 108+ – concepiscono il diritto di accesso come elemento strutturale della legittimità del trattamento¹⁷.

Nel diritto dell'Unione europea, la Direttiva (UE) 2016/680 (LED) e il GDPR sanciscono il diritto di accesso come componente del diritto fondamentale alla protezione dei dati, pur ammettendo

restrizioni per finalità investigative. La dottrina ha evidenziato come, nei contesti investigativi, tali diritti siano spesso svuotati di contenuto a causa di clausole di eccezione e di accesso mediato tramite autorità di controllo. Su questo terreno, la giurisprudenza di Strasburgo offre indicazioni decisive: in *Zakharov c. Russia*, *Szabó e Vissy c. Ungheria* e *Big Brother Watch e altri c. Regno Unito*, la Corte ha sottolineato il nesso tra notifica, diritto di accesso e rimedi effettivi – l'assenza di qualsiasi forma di informazione rende impossibile un controllo giurisdizionale *ex post* – e ha chiarito che la mera esistenza formale di un'autorità di controllo non è sufficiente se l'interessato non può contestare concretamente la liceità del trattamento¹⁸.

Sul piano metodologico, questi sviluppi giurisprudenziali rilevano direttamente per la costruzione della variabile A nell'IICR. La dottrina di informatica giuridica ha da tempo riconosciuto questa tensione: Sartor ha evidenziato come i sistemi informatici di supporto alla decisione giuridica richiedano criteri di valutazione specifici, distinti da quelli elaborati per le tecnologie tradizionali; Hildebrandt ha sottolineato come le "smart technologies" alterino strutturalmente il rapporto tra individuo e Stato; Simoncini ha dimostrato come i sistemi AI possano configurare una "incostituzionalità algoritmica" quando operano in assenza di trasparenza e controllo umano effettivo; Koops e Leenes hanno messo in luce i limiti del paradigma della privacy by design in contesti di law enforcement; Mantelero ed Esposito hanno proposto una metodologia di valutazione dell'impatto sui diritti umani (HRIA) nei sistemi AI ad uso intensivo di dati; Manes ha analizzato il rapporto tra oracolo algoritmico e garanzie processuali; Santosuosso e

16. Corte EDU, *Yonchev c. Bulgaria*, ric. n. 12504/09, 7 dicembre 2017; Corte EDU, *Joanna Szulc c. Polonia*, ric. n. 43932/08, 13 novembre 2012; Corte EDU, *Jarnea c. Romania*, ric. n. 41838/05, 19 luglio 2011; Corte EDU, *Haralambie c. Romania*, ric. n. 21737/03, 27 ottobre 2009; Corte EDU, Grande Camera, *Rotaru c. Romania*, ric. n. 28341/95, 4 maggio 2000; Corte EDU, *Gaskin c. Regno Unito*, ric. n. 10454/83, 7 luglio 1989.

17. Consiglio d'Europa, Comitato dei Ministri, Raccomandazione n. R (87) 15 relativa all'uso dei dati personali nel settore della polizia, 17 settembre 1987; Consiglio d'Europa, Protocollo di emendamento della Convenzione sulla protezione delle persone rispetto al trattamento automatizzato di dati a carattere personale (CETS n. 223, Convenzione 108+), 10 ottobre 2018; Consiglio d'Europa, Comitato consultivo della Convenzione 108, *Practical Guide on the Use of Personal Data in the Police Sector*, T-PD(2018)01, 16 febbraio 2018.

18. Corte EDU, Grande Camera, *Big Brother Watch e altri c. Regno Unito*, ricc. nn. 58170/13, 62322/14 e 24960/15, 25 maggio 2021; Corte EDU, Grande Camera, *Centrum för rättvisa c. Svezia*, ric. n. 35252/08, 25 maggio 2021; Corte EDU, *Szabó e Vissy c. Ungheria*, ric. n. 37138/14, 12 gennaio 2016; Corte EDU, Grande Camera, *Roman Zakharov c. Russia*, ric. n. 47143/06, 4 dicembre 2015; Corte EDU, *Dalea c. Francia*, ric. n. 964/07, 2 febbraio 2010.

Sartor hanno individuato le condizioni di compatibilità della giustizia predittiva con i principi del giusto processo¹⁹. Tutti questi contributi convergono nel ritenere che l'affidabilità tecnica del sistema informatico impiegato non sia un requisito meramente ingegneristico, ma una preconditione giuridica per la legittimità dell'intervento investigativo. Un sistema informatico opaco o inaffidabile abbassa il valore di A, facendo lievitare l'IICR oltre la soglia di ammissibilità costituzionale: nega all'individuo – e agli organi di controllo – la possibilità di contestare concretamente la misura, replicando sul piano algoritmico la stessa lesione che la Corte ha riconosciuto nei casi di accesso negato ai fascicoli statali.

5. Trasparenza e controllo indipendente

I contesti di sorveglianza segreta e di cooperazione investigativa transnazionale mettono in luce i limiti strutturali di un modello di tutela fondato esclusivamente sull'informazione diretta dell'interessato²⁰. In tali ambiti, la trasparenza non può essere intesa come requisito assoluto, bensì come variabile funzionale all'effettività del controllo sulle attività statali, suscettibile di restrizioni purché compensate da garanzie alternative che preservino il nucleo essenziale dei diritti protetti dall'art. 8 CEDU²¹.

Sin da *Klass c. Germania* e *Leander c. Svezia*, la Corte ha ritenuto che l'assenza di notifica individuale possa essere compatibile con la Convenzione solo se l'ordinamento prevede forme di supervisione esterna dotate di indipendenza e poteri effettivi di verifica della legalità delle misure adottate²². In *Zakharov c. Russia* e *Szabó e Vissy c. Ungheria*, la Corte ha precisato che il divieto generalizzato di

notifica rende impossibile un controllo giurisdizionale concreto²³. *Big Brother Watch e altri c. Regno Unito* rafforza questo orientamento, sottolineando la necessità di un sistema articolato di garanzie procedurali che consenta almeno una contestazione *ex post* delle ingerenze. Nel diritto dell'Unione europea, la Direttiva (UE) 2016/680 (Law Enforcement Directive, LED) attribuisce alle autorità di controllo un ruolo centrale, con poteri effettivi di verifica sostanziale²⁴.

Nel contesto delle investigazioni supportate da sistemi informatici di intelligenza artificiale, queste indicazioni si traducono in un requisito specifico per la variabile A: un sistema che non consente supervisione indipendente e verificabile – perché la sua architettura o i suoi parametri decisionali non sono accessibili agli organi di controllo – tende strutturalmente a ridurre il proprio valore di affidabilità. Ciò non per una valutazione astratta, ma perché la giurisprudenza sul controllo indipendente – da *Klass* a *Big Brother Watch* – ha chiarito che l'effettività della supervisione deve essere verificata caso per caso in relazione ai poteri reali di cui l'organo di controllo dispone²⁵. Il presente contributo propone che questa verifica sia incorporata nella valutazione della variabile A, piuttosto che affidata a un giudizio separato e successivo.

6. Giurisprudenza della Corte EDU e non-disclosure di dati personali

La giurisprudenza della Corte ha chiarito che limitazioni al diritto di accesso motivabili per ragioni di sicurezza nazionale, investigazione penale o prevenzione del crimine organizzato devono essere necessarie, proporzionate e accompagnate da meccanismi di supervisione indipendente e rimedi

19. SANTOSUOSSO–SARTOR 2022; SARTOR 2022-B; MANTELERO–ESPOSITO 2021; MANES 2020; SIMONCINI 2019; HILDEBRANDT 2015; KOOPS–LEENES 2014.

20. CARAMASCHI 2025; SIMONCINI 2019.

21. MARTIRE 2024.

22. Corte EDU, *Leander c. Svezia*, ric. n. 9248/81, 26 marzo 1987; Corte EDU, *Klass e altri c. Germania*, ric. n. 5029/71, 6 settembre 1978.

23. Corte EDU, *Grande Camera, Big Brother Watch e altri c. Regno Unito*, ricc. nn. 58170/13, 62322/14 e 24960/15, 25 maggio 2021; Corte EDU, *Szabó e Vissy c. Ungheria*, ric. n. 37138/14, 12 gennaio 2016; Corte EDU, *Grande Camera, Roman Zakharov c. Russia*, ric. n. 47143/06, 4 dicembre 2015.

24. FALLETTA–MARSANO 2024; PIETRANGELO 2024.

25. Corte EDU, *Grande Camera, Big Brother Watch e altri c. Regno Unito*, ricc. nn. 58170/13, 62322/14 e 24960/15, 25 maggio 2021; Mantelero-Esposito 2021; Corte EDU, *Klass e altri c. Germania*, ric. n. 5029/71, 6 settembre 1978.

effettivi. In *Szulc c. Polonia* e *Jarnea c. Romania*, il diritto di accesso è stato esteso anche ai dati conservati a fini storici o di sorveglianza politica: un orientamento che, ai fini del modello proposto, implica che la variabile *A* debba tenere conto non solo dei dati direttamente impiegati nell'indagine, ma anche di quelli conservati in archivi secondari e potenzialmente accessibili ai sistemi informatici di analisi.

In scenari di sorveglianza digitale, come evidenziato nei casi *Zakharov c. Russia*, *Szabó e Vissy c. Ungheria* ed *Ekimdzhiiev e altri c. Bulgaria*²⁶, la Corte sottolinea che divieti assoluti di notifica possono compromettere l'efficacia dei rimedi e impedire il controllo giurisdizionale. In tali contesti, l'opacità dei sistemi informatici investigativi accentua la necessità di garanzie multilivello che integrino oversight indipendente, auditing, notifiche *post-facto* e possibilità di ricorso.

Due indicazioni emergono con chiarezza e orientano direttamente la costruzione delle variabili del modello: da un lato, la notifica *post-facto* costituisce una garanzia minima non comprimibile – la sua esclusione strutturale da parte di un sistema informatico investigativo abbassa il valore di *A* perché pregiudica la possibilità di rimedi effettivi; dall'altro, la gravità della minaccia non può di per sé legittimare qualsiasi misura, ma deve essere posta in rapporto proporzionato con l'intensità dell'ingerenza – principio che trova espressione nella formula *IICR* nell'impossibilità che un alto valore di *M* compensi un basso valore di *A*. Queste indicazioni giurisprudenziali costituiscono il presupposto normativo delle sezioni successive, che analizzano come le specificità tecniche dei sistemi AI – opacità algoritmica, rischio cibernetico, profilazione predittiva – modifichino concretamente i parametri di valutazione della proporzionalità.

7. Opacità algoritmica e crittografia nelle investigazioni AI-based

Il presente paragrafo esamina due profili tecnici distinti che incidono, con meccanismi differenti,

sulla valutazione della proporzionalità nelle investigazioni basate su sistemi informatici di intelligenza artificiale: l'*opacità algoritmica* e i problemi connessi alla *crittografia* e alla decrittazione. Si tratta di fenomeni tecnologicamente differenti – l'opacità algoritmica riguarda l'impossibilità di ricostruire il percorso decisionale di un sistema informatico di intelligenza artificiale; la crittografia protegge i dati in transito o a riposo attraverso meccanismi matematici di cifratura – ma accomunati da un medesimo effetto giuridico: entrambi limitano la capacità dell'interessato e degli organi di controllo di verificare le modalità con cui i dati personali vengono trattati nell'ambito delle indagini²⁷. È questa convergenza sul piano degli effetti che giustifica la trattazione congiunta in relazione alla costruzione della variabile *A* del modello.

L'opacità algoritmica – ossia l'impossibilità di accedere e comprendere i parametri decisionali di un sistema informatico di analisi – costituisce il principale fattore di riduzione del valore di *A*. Quando le autorità competenti non possono accedere alla documentazione tecnica del sistema – architettura, dati di addestramento, parametri decisionali – la verifica dell'affidabilità dei risultati prodotti diventa impossibile, replicando sul piano tecnico la stessa lesione del diritto di controllo che la Corte ha censurato nei casi di accesso negato ai fascicoli statali²⁸. Il Regolamento sull'intelligenza artificiale (AI Act) risponde in parte a questa esigenza imponendo obblighi di trasparenza e documentazione per i sistemi AI ad alto rischio, inclusi quelli impiegati in contesti di law enforcement. Tuttavia, i requisiti di trasparenza algoritmica non equivalgono automaticamente a supervisione effettiva: occorre che le autorità di controllo dispongano anche dei mezzi tecnici e delle competenze per esercitare quella supervisione.

Quanto alla crittografia, la richiesta di accesso forzato a sistemi criptati o di backdoor tecnologiche impone una valutazione specifica di proporzionalità²⁹. Ogni operazione di accesso forzato o decrittazione disposta dall'autorità giudiziaria deve

26. Corte EDU, *Ekimdzhiiev e altri c. Bulgaria*, ric. n. 70078/12, 11 gennaio 2022

27. FINOCCHIARO 2024; SARTOR 2022-A; KOOPS-LEENES 2014.

28. Cfr. Corte EDU, *Haralambie c. Romania*, ric. n. 21737/03, 27 ottobre 2009; Corte EDU, *Grande Camera, Rotaru c. Romania*, ric. n. 28341/95, 4 maggio 2000; Corte EDU, *Gaskin c. Regno Unito*, ric. n. 10454/83, 7 luglio 1989.

Sugli obblighi di documentazione tecnica dei sistemi ad alto rischio, Regolamento (UE) 2024/1689, artt. 11-13.

29. DANEZIS-DOMINGO-FERRER-HANSEN et al. 2014.

essere documentata, proporzionata alla gravità del reato e sottoposta a supervisione giudiziaria preventiva e ad auditing tecnico successivo. Il problema giuridicamente rilevante non è la crittografia in sé – che costituisce un presidio legittimo della sicurezza dei dati – bensì la risposta istituzionale alla crittografia: un sistema di decrittazione che introduce vulnerabilità sistemiche per consentire l'accesso investigativo abbassa strutturalmente il valore di *A*, perché aumenta il rischio di esposizione dei dati a soggetti non autorizzati e compromette l'integrità probatoria dei risultati³⁰.

I due profili tecnici esaminati – opacità algoritmica e vulnerabilità crittografiche – sono stati considerati come sotto-indicatori della variabile *A* nel dataset esplorativo costruito nel contesto della ricerca. L'esperienza di osservazione presso il Justice Digital Center di Wrocław ha suggerito che entrambi i fattori incidono significativamente sul valore di *A* e, per questa via, sull'IICR: un sistema informatico investigativo che presenti elevata opacità algoritmica o che richieda decrittazione sistematica senza adeguate garanzie tende a produrre valori di intrusività costituzionalmente rilevante incompatibili con il principio di proporzionalità, indipendentemente dalla gravità del reato perseguito. Si tratta tuttavia di indicazioni preliminari che richiedono ulteriori verifiche empiriche.

8. Cybersecurity e rischio investigativo

L'integrazione di sistemi informatici di intelligenza artificiale nelle investigazioni transnazionali introduce rischi concreti legati alla cybersecurity che incidono sulla proporzionalità delle misure adottate³¹. I dati trattati dalle autorità investigative, soprattutto in contesti di cooperazione internazionale antimafia e cyber investigation, includono informazioni sensibili la cui compromissione può avere effetti devastanti sia sui diritti fondamentali degli interessati sia sull'efficacia stessa delle indagini. La cybersecurity è qui intesa come l'insieme delle misure tecniche e organizzative volte a proteggere sistemi, reti e dati da attacchi, accessi non autorizzati e malfunzionamenti.

Le vulnerabilità informatiche possono derivare da molteplici fonti: errori nella configurazione dei sistemi, insufficiente aggiornamento dei protocolli di sicurezza, esposizione di database condivisi tra Stati membri con standard eterogenei, attacchi mirati da parte di soggetti criminali organizzati. Nella prassi investigativa transnazionale, i sistemi concretamente impiegati – piattaforme di analisi predittiva come Palantir Gotham, il sistema SIENA di Europol per lo scambio di informazioni operative, strumenti di riconoscimento facciale come SARI, software di predictive policing – presentano specifici profili di rischio cibernetico. L'interoperabilità tra piattaforme con architetture diverse, la conservazione di dataset di addestramento provenienti da ordinamenti con standard differenti e la vulnerabilità delle interfacce di connessione tra sistemi nazionali sono i principali vettori di esposizione identificati nella letteratura³².

Questi rischi si collegano ai profili esaminati nel paragrafo 7 in modo specifico: mentre l'opacità algoritmica oscura il percorso decisionale del sistema e le vulnerabilità crittografiche espongono i dati in transito, le vulnerabilità di cybersecurity espongono i medesimi dati a rischi esterni. Il rischio cibernetico costituisce pertanto un fattore autonomo nella valutazione della proporzionalità della misura: ogni sistema investigativo basato su tecnologie informatiche deve essere valutato non solo in termini di necessità rispetto all'obiettivo investigativo, ma anche considerando il livello di esposizione del sistema a compromissioni o manipolazioni esterne³³. Maggiore è tale esposizione, maggiore è l'intrusività reale della misura, perché le informazioni raccolte possono raggiungere soggetti non autorizzati, generando danni collaterali sui diritti fondamentali degli interessati e potenzialmente di terzi.

Sul piano della costruzione della variabile *A*, il rischio cibernetico agisce con una doppia incidenza: abbassa direttamente il valore di *A* – riducendo l'affidabilità e la resilienza del sistema – e può incrementare *S* qualora la compromissione dei dati produca danni istituzionali o socio-economici diffusi. Questa doppia incidenza spiega perché, nel dataset

30. MANTELERO-ESPOSITO 2021; DANEZIS-DOMINGO-FERRER-HANSEN et al. 2014.

31. POLLICINO 2025; PIETRANGELO 2024.

32. FIORINELLI 2025; MANTELERO-ESPOSITO 2021; DANEZIS-DOMINGO-FERRER-HANSEN et al. 2014.

33. POLLICINO 2025.

esplorativo, scenari con elevato rischio cibernetico generassero valori di *IICR* particolarmente elevati anche in presenza di minacce criminali di gravità media. Per mitigare questi rischi, il modello prevede l'integrazione di: valutazioni sistematiche del rischio (*risk assessment*) per ogni sistema informatico di analisi; standard di sicurezza uniformi tra Stati membri compatibili con GDPR e LED; procedure di auditing e supervisione indipendente con tracciamento degli accessi; misure di resilienza e recovery. La valutazione sistematica del rischio cibernetico – attraverso threat modeling, audit di vulnerabilità e standard ENISA – è una componente non opzionale del giudizio di proporzionalità nelle indagini AI-based.

9. Profilazione predittiva e intrusività anticipatoria

La profilazione predittiva basata su sistemi informatici di intelligenza artificiale introduce un livello di *intrusività anticipatoria* che costituisce un'ulteriore estensione delle problematiche esaminate nei paragrafi 7 e 8. Con questa espressione si intende qui l'interferenza sui diritti fondamentali che si produce prima che emerga un sospetto concreto, attraverso l'elaborazione di dati aggregati da parte di software di analisi predittiva che generano inferenze sul comportamento futuro di individui o gruppi³⁴. Questo comporta rischi di ingiusta stigmatizzazione, interferenze preventive sui diritti fondamentali e vulnerabilità sistemiche in contesti transnazionali, potenzialmente amplificati da criticità di sicurezza informatica o da bias nei sistemi impiegati³⁵.

Dal punto di vista giuridico, la giurisprudenza della CEDU fornisce strumenti interpretativi rilevanti. Nei casi *Segerstedt-Wiberg c. Svezia* (2006) e *Zakharov c. Russia* (2015), la Corte ha sottolineato che le misure di sorveglianza devono essere valutate non solo per la loro legittimità formale, ma anche in relazione all'effettiva intensità dell'intrusione e alla probabilità di errori o abusi³⁶. Trasponendo questi principi alla profilazione predittiva,

l'affidabilità dei modelli algoritmici, la gestione del bias e la robustezza del sistema informatico sono determinanti per stabilire la proporzionalità della misura: software con elevata probabilità di falsi positivi o bias sistemico aumentano l'intrusività reale e richiedono garanzie aggiuntive di supervisione, revisione e limitazione del trattamento³⁷.

In contesti di cooperazione internazionale antimafia, i sistemi effettivamente impiegati o sperimentati in ambito europeo includono: il sistema iBorderCtrl (progetto pilota Ue per il rilevamento predittivo di comportamenti sospetti alle frontiere), i moduli di analisi delle reti criminali nella piattaforma SIENA di Europol, i sistemi di analisi dei flussi finanziari illeciti nell'ambito di Eurojust e delle Joint Investigation Teams, strumenti di link analysis come Analysts' Notebook di IBM. L'osservazione condotta presso il Justice Digital Center di Wrocław ha suggerito che la variabile *A* risulta particolarmente sensibile alla qualità dei dati di addestramento provenienti da ordinamenti con diversi standard probatori³⁸. Si tratta di un'indicazione che richiede verifica su campioni più ampi, ma che segnala un'area di criticità metodologica rilevante.

La profilazione predittiva introduce inoltre una dimensione temporale della proporzionalità che ha richiesto una scelta metodologica specifica. Nei modelli tradizionali, la variabile *M* – gravità della minaccia – è definita *ex post* rispetto a un reato già commesso o a un sospetto fondato. Nella profilazione predittiva, l'intervento investigativo precede la materializzazione della minaccia, e *M* deve essere stimata sulla base di inferenze algoritmiche. Questo introduce un'asimmetria strutturale: quanto meno affidabile è il sistema predittivo (basso *A*), tanto più incerta è la stima di *M*, e tanto più elevato diventa il valore di *IICR*. Nel dataset esplorativo, gli scenari di profilazione predittiva sono stati trattati come categoria distinta, con soglie più restrittive: l'ingerenza anticipatoria appare proporzionata solo quando il valore di *A* è elevato e documentato, perché solo un sistema informatico ad alta

34. SANTOSUOSSO-SARTOR 2022; MANES 2020.

35. FIORINELLI 2025.

36. Corte EDU, Grande Camera, *Roman Zakharov c. Russia*, ric. n. 47143/06, 4 dicembre 2015; Corte EDU, *Segerstedt-Wiberg e altri c. Svezia*, ric. n. 62332/00, 6 giugno 2006.

37. MANTELERO-ESPOSITO 2021.

38. GUERRA 2025.

affidabilità tecnica e probatoria può giustificare un'interferenza preventiva sui diritti fondamentali. Anche qui, tuttavia, si tratta di ipotesi da sottoporre a verifica empirica sistematica.

Le regole operative formulate al paragrafo 3.3 trovano in questo regime la loro applicazione più rilevante. Negli scenari predittivi, la valutazione si articola in due stadi: in un primo stadio va accertata *A* – su base informativa indipendente dal caso concreto, secondo la validazione *ex ante* descritta al paragrafo 3.1 – verificando il superamento della soglia rafforzata *A**; solo nel secondo stadio, e nei soli scenari che superano tale soglia, la stima algoritmica di *M* è ammessa nel calcolo dell'indice, nel suo valore prudentiale e subordinatamente alla presenza di riscontri investigativi esterni al sistema. Ne discende un'asimmetria voluta: quando *A* è basso, la stima di *M* non viene semplicemente scontata, ma diviene inutilizzabile – il sistema non può autolegittimarsi attraverso la gravità della minaccia che esso stesso ha stimato. È in questo senso che l'interdipendenza tra le due variabili, riconosciuta come nodo strutturale del modello, viene trattata non come vizio da occultare ma come vincolo di ammissibilità da rendere esplicito.

10. Affidabilità probatoria come sotto-indicatore autonomo

L'analisi condotta nei paragrafi 7-9 mette in luce come l'intrusività delle investigazioni transnazionali basate su sistemi informatici di intelligenza artificiale non si misuri più solo in termini di accesso ai dati o di intensità della sorveglianza, ma anche attraverso la *qualità epistemica dell'evidenza prodotta*³⁹. Un dato impreciso, una predizione errata o un sistema informatico non validato può avere effetti sproporzionati sui diritti fondamentali, generando interferenze preventive invisibili e anticipatorie⁴⁰. L'affidabilità probatoria diventa così un sotto-indicatore autonomo della variabile *A*.

La giurisprudenza della CEDU offre indicazioni operative significative: in *Gaskin c. Regno Unito* (1989) e *Rotaru c. Romania* (2000), la Corte ha evidenziato come la raccolta e il trattamento dei dati personali debbano essere sufficientemente affidabili e documentati per giustificare l'interferenza sul diritto alla vita privata⁴¹. I software di analisi predittiva utilizzati nella cooperazione internazionale antimafia o nelle cyber investigation operano su dati eterogenei, provenienti da Stati membri con standard differenti, aumentando il rischio di errori sistemici⁴². Sistemi privi di controlli adeguati possono generare falsi positivi, segnalare erroneamente individui come sospetti o influenzare decisioni investigative senza adeguata verifica⁴³.

Indicatori pratici per la valutazione dell'affidabilità probatoria – e dunque per la determinazione del valore di *A* – includono:

1. precisione e recall dei software predittivi, per stimare falsi positivi e negativi⁴⁴;
2. documentazione del flusso dei dati, per verificare provenienza, integrità e aggiornamento delle informazioni;
3. audit e revisione indipendente, inclusi test retrospettivi sulla correttezza delle predizioni⁴⁵;
4. allineamento alle normative transnazionali – GDPR, LED, Convention 108+, CM(2018)2, Raccomandazione R(87)15 – che garantiscano fondamento legale e trattamento proporzionato dei dati.

L'analisi condotta in questo paragrafo chiude il fondamento argomentativo dell'IICR: la qualità epistemica dell'evidenza prodotta dal sistema informatico – misurata attraverso i quattro indicatori indicati – è parte integrante dell'affidabilità tecnica e probatoria che la variabile *A* è chiamata a esprimere. Nelle investigazioni transnazionali basate su AI, un sistema inaffidabile sul piano probatorio non è semplicemente meno utile: produce un'intrusività reale anche quando l'accesso ai dati

39. MANTELERO-ESPOSITO 2021.

40. DANEZIS-DOMINGO-FERRER-HANSEN et al. 2014.

41. Corte EDU, Grande Camera, *Rotaru c. Romania*, ric. n. 28341/95, 4 maggio 2000; Corte EDU, *Gaskin c. Regno Unito*, ric. n. 10454/83, 7 luglio 1989.

42. PIETROPAOLI 2024.

43. FIORINELLI 2025.

44. MANTELERO-ESPOSITO 2021; DANEZIS-DOMINGO-FERRER-HANSEN et al. 2014.

45. MANTELERO-ESPOSITO 2021; CEPEJ 2018.

è formalmente legittimo, perché condiziona decisioni umane sulla base di evidenze non verificabili.

11. Dal framework PEGA a PEGA 2: continuità e aggiornamento

L'analisi sviluppata nei paragrafi 4-10 consente ora di precisare il rapporto tra il framework PEGA e la proposta PEGA 2, chiarendo tanto gli elementi di continuità quanto quelli di aggiornamento metodologico.

Il framework PEGA ha il merito storico di aver introdotto, in un contesto istituzionale concreto, criteri operativi per la valutazione della proporzionalità delle misure di sorveglianza. I suoi cinque criteri qualitativi – autorizzazione *ex ante*, proporzionalità e necessità, supervisione indipendente, notifica, cancellazione dei dati non pertinenti – restano pienamente validi e PEGA 2 li incorpora integralmente. La continuità con il framework originario è un elemento strutturale della proposta, non una concessione: i principi convenzionali su cui si fonda PEGA non sono superati dall'evoluzione tecnologica, ma richiedono di essere integrati con parametri capaci di intercettare le specificità dei sistemi informatici di intelligenza artificiale.

L'aggiornamento metodologico proposto con PEGA 2 si articola su tre livelli. Il primo livello riguarda la dimensione tecnica: il framework PEGA non considerava l'affidabilità tecnica e probatoria dei sistemi informatici, il rischio di bias algoritmico, la resilienza cibernetica e l'opacità algoritmica. PEGA 2 incorpora questi elementi nella variabile A, traducendo in parametri di valutazione della proporzionalità ciò che nei sistemi di intelligenza artificiale costituisce una preconditione giuridica – non solo ingegneristica – della legittimità dell'intervento investigativo. Il secondo livello riguarda la dimensione socio-economica: le misure investigative producono effetti che si estendono oltre la sfera del singolo interessato, incidendo su istituzioni, mercati e coesione sociale. La variabile S incorpora questa dimensione, che il framework PEGA non contemplava. Il terzo livello riguarda l'operatività: mentre il framework PEGA offre criteri qualitativi di conformità, l'IICR propone un quadro strutturato di valutazione che impone di esplicitare e ponderare le variabili rilevanti, rendendo trasparente il ragionamento proporzionalistico.

Il confronto tra PEGA e PEGA 2 può essere sintetizzato come segue. Il framework PEGA risponde alla domanda: *la misura è proceduralmente legittima?* PEGA 2 aggiunge la domanda: *quanto è realmente intrusiva, considerata la qualità tecnica del sistema informatico impiegato, la sua opacità, la sua vulnerabilità ai rischi cibernetici e il suo impatto sociale?* La prima domanda è necessaria ma non sufficiente. La seconda è quella che il diritto fatica ancora a formulare in modo sistematico e che PEGA 2 cerca di rendere operativa.

Va sottolineato che PEGA 2, nella sua attuale formulazione, costituisce una proposta preliminare che richiede ulteriori verifiche. Il dataset esplorativo costruito nel contesto della ricerca – cinquanta scenari investigativi reali e simulati osservati presso il Justice Digital Center di Wrocław – ha consentito di testare la tenuta delle variabili e di orientare la definizione delle soglie, ma non può essere considerato sufficiente per una validazione sistematica del modello⁴⁶. Le soglie proposte sono indicative e devono essere verificate su campioni più ampi, in contesti operativi differenti e con il contributo di competenze multidisciplinari. L'IICR è uno strumento analitico aperto, non un algoritmo decisionale chiuso.

Per dare concretezza a quanto precede, è utile esplicitare lo schema di codifica che il modello richiede per la costruzione di dataset di scenari, e che ha orientato la costruzione del dataset esplorativo. Ai fini della replicabilità, ciascuno scenario va codificato secondo campi omogenei: la tipologia investigativa (indagine *ex post*, monitoraggio in corso, profilazione predittiva); la categoria di sistema informatico impiegato (analisi predittiva, link analysis, riconoscimento facciale, piattaforme di scambio informativo transnazionale); la fattispecie di reato perseguita; i valori assegnati a M, A e S secondo le griglie del paragrafo 3.1, con la relativa motivazione; il valore risultante dell'IICR; l'esito valutativo. Nelle esemplificazioni che seguono si assumono, a titolo puramente illustrativo, una soglia di ammissibilità dell'indice pari a 2 e una soglia rafforzata di validazione A* pari a 0,7 per il regime predittivo: valori indicativi, la cui calibrazione è rimessa alle verifiche empiriche future.

46. Cfr. *supra*, nota 7.

Tre scenari-tipo, costruiti con valori puramente ipotetici a fini illustrativi, mostrano il funzionamento del modello.

Nel primo, un'indagine *ex post* sui flussi finanziari di un'organizzazione criminale transnazionale impiega una piattaforma di link analysis corredata di documentazione tecnica completa e sottoposta ad audit indipendente: la gravità della minaccia è elevata e ancorata ai cataloghi normativi ($M = 0,9$), il sistema è affidabile ($A = 0,8$), l'impatto socio-economico contenuto ($S = 0,2$); l'indice risultante – circa 1,4 – resta al di sotto della soglia indicativa e la misura appare proporzionata.

Nel secondo, l'impiego di un sistema di riconoscimento facciale su larga scala per un reato di media gravità, in assenza di documentazione sul tasso di falsi positivi e di audit ($M = 0,5$; $A = 0,3$; $S = 0,4$), produce un indice di circa 2,8: la sproporzione non deriva dalla gravità del reato, ma dall'inaffidabilità del sistema e dall'ampiezza dell'impatto sociale – esattamente la dimensione che i criteri qualitativi del framework PEGA non intercettano.

Nel terzo, uno scenario di profilazione predittiva alle frontiere, la gravità della minaccia è stimata dallo stesso sistema impiegato: la validazione *ex ante* restituisce $A = 0,4$, inferiore alla soglia rafforzata, e la misura risulta inammissibile a prescindere dal calcolo dell'indice, secondo la regola del paragrafo 3.3.

I tre scenari illustrano la capacità dell'indice di discriminare situazioni che un controllo meramente procedurale tratterebbe in modo indifferenziato; resta ferma la necessità, già segnalata, di una validazione su campioni più ampi.

12. Conclusioni

Il presente contributo ha analizzato i limiti del framework PEGA quando applicato al più ampio scenario delle investigazioni basate su sistemi informatici di intelligenza artificiale e ha proposto PEGA 2 come aggiornamento metodologico che ne conserva l'impianto convenzionale integrandolo con parametri tecnici, socio-economici e giuridici. L'elemento centrale della proposta è l'IICR, uno strumento analitico che traduce il principio di proporzionalità nella relazione tra gravità della minaccia criminale (M), affidabilità tecnica e probatoria

del sistema informatico impiegato (A) e impatto socio-economico della misura (S)⁴⁷. Rispetto a una mera enunciazione di principio, il contributo ha indicato criteri operativi per la determinazione dei valori delle variabili (par. 3.1), ha chiarito la natura – predeterminata, semi-strutturata o valutativa – dei parametri impiegati (par. 3.2) e ha affrontato l'interdipendenza tra M e A nel regime predittivo, incorporandola nel modello come vincolo di ammissibilità (parr. 3.3 e 9).

L'analisi giurisprudenziale sviluppata nei paragrafi 4-6 ha mostrato come la giurisprudenza della Corte europea dei diritti dell'uomo in materia di accesso ai dati, trasparenza e non-disclosure fornisca già gli elementi per una valutazione sostanziale – non solo procedurale – dell'intrusività delle misure investigative. I paragrafi 7-10 hanno tradotto questi elementi in criteri operativi per la costruzione della variabile A : l'opacità algoritmica, le vulnerabilità crittografiche, il rischio cibernetico e l'affidabilità probatoria sono fattori che determinano concretamente quanto sia intrusiva una misura investigativa AI-based dal punto di vista costituzionale.

Il modello proposto consente di affrontare alcune delle principali sfide che le investigazioni AI-based pongono al principio di proporzionalità: prevenire interventi eccessivamente invasivi rispetto al rischio concreto; integrare valutazioni tecnico-scientifiche, giuridiche e socio-economiche nel giudizio di proporzionalità; armonizzare le attività investigative transnazionali, riducendo i rischi derivanti da disallineamenti normativi o vulnerabilità cibernetiche; rendere trasparente il ragionamento proporzionalistico, tanto per le autorità giudiziarie quanto per le agenzie investigative⁴⁸.

Nella sua attuale formulazione, PEGA 2 è una proposta metodologica che richiede ulteriori sviluppi. In prospettiva futura, il modello potrebbe essere arricchito mediante: l'integrazione di metriche avanzate di affidabilità algoritmica e spiegabilità dei sistemi informatici; la validazione su campioni più ampi di scenari investigativi reali; lo sviluppo di framework comuni per la cooperazione internazionale basati su standard condivisi di cybersecurity; la sperimentazione in contesti operativi concreti per affinare le soglie proposte. L'obiettivo non è produrre un sistema chiuso, ma

47. Raccomandazione del Parlamento europeo B9-0260/2023 del 15 giugno 2023, §§ 30-32; Relazione A9-0189/2023.

48. MANTELERO-ESPOSITO 2021; DANEZIS-DOMINGO-FERRER-HANSEN et al. 2014.

offrire all'informatica giuridica un quadro concettuale che traduca la complessità tecnica dei sistemi di intelligenza artificiale in categorie giuridicamente maneggevoli⁴⁹.

Riferimenti bibliografici

- R. BRIGHI (a cura di) (2022), *Nuove questioni di informatica forense*, Aracne, 2022
- R. BRIGHI, M. PALMIRANI, M.E. SÁNCHEZ JORDÁN (a cura di) (2018), *Informatica giuridica e informatica forense al servizio della società della conoscenza. Scritti in onore di Cesare Maioli*, Gioacchino Onorati, 2018
- L. CAMALDO (2025), *Intelligenza artificiale e investigazione penale predittiva*, in F. Basile, M. Biasi, L. Camaldo, G. Caneschi, B. Fragasso, D. Milani (a cura di), "Intelligenza artificiale. Diritto, giustizia, economia ed etica", Giappichelli, 2025
- O. CARAMASCHI (2025), *Il costituzionalismo al cospetto dell'intelligenza artificiale: nuove sfide, quali soluzioni?*, in "Rivista italiana di informatica e diritto", vol. 7, 2025, n. 1
- CEPEJ – COMMISSIONE EUROPEA PER L'EFFICIENZA DELLA GIUSTIZIA (2018), *Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*, adottata il 3-4 dicembre 2018, Consiglio d'Europa, 2018
- S. COGNETTI (2011), *Principio di proporzionalità. Profili di teoria generale e di analisi sistematica*, Giappichelli, 2011
- G. DANEZIS, J. DOMINGO-FERRER, M. HANSEN, J.-H. HOEPFMAN, D. LE MÉTAYER, R. TIRTEA, S. SCHIFFNER (2014), *Privacy and data protection by design – from policy to engineering*, ENISA, dicembre 2014 (disponibile anche in arXiv:1501.03726, 2015)
- P. FALLETTA, A. MARSANO (2024), *Intelligenza artificiale e protezione dei dati personali: il rapporto tra regolamento europeo sull'intelligenza artificiale e GDPR*, in "Rivista italiana di informatica e diritto", vol. 6, 2024, n. 1
- E. FAMELI (2025), *La nuova intelligenza artificiale e il problema della sua regolazione giuridica. La strategia europea e il lungo iter legislativo dell'AI Act*, in "Rivista italiana di informatica e diritto", vol. 7, 2025, n. 1
- G. FINOCCHIARO (2024), *Diritto dell'intelligenza artificiale*, Zanichelli, 2024
- G. FIORINELLI (2025), *La recente "disciplina penale" del cybercrime tra armonizzazione internazionale e interventi nazionali: convergenze parallele?*, in "Rivista italiana di informatica e diritto", vol. 7, 2025, n. 2
- M. GUERRA (2025), *Le prospettive di collaborazione tra Autorità indipendenti nell'era digitale*, in "Rivista italiana di informatica e diritto", vol. 7, 2025, n. 1
- M. HILDEBRANDT (2015), *Smart technologies and the end(s) of law. Novel entanglements of law and technology*, Edward Elgar, 2015
- B.-J. KOOPS, R. LEENES (2014), *Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data-protection law*, in "International Review of Law, Computers & Technology", vol. 28, 2014, n. 2
- V. MANES (2020), *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in U. Ruffolo (a cura di), "Intelligenza artificiale. Il diritto, i diritti, l'etica", Giuffrè Francis Lefebvre, 2020

49. L'autrice esprime la propria gratitudine al team del Justice Digital Center di Wrocław per il supporto scientifico e l'accoglienza forniti durante il periodo di visiting research.

- A. MANTELERO, M.S. ESPOSITO (2021), *An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems*, in "Computer Law & Security Review", vol. 41, 2021
- D. MARTIRE (2024), *Human in the loop. L'essere umano come fattore condizionante della – o condizionato dalla – intelligenza artificiale*, in "Rivista italiana di informatica e diritto", vol. 6, 2024, n. 2
- U. PAGALLO (2013), *The laws of robots. Crimes, contracts, and torts*, Springer, 2013
- M. PALMIRANI, S. SAPIENZA (a cura di) (2022), *La trasformazione digitale della giustizia nel dialogo tra discipline. Diritti e Intelligenza Artificiale*, Giuffrè Francis Lefebvre, 2022
- M. PIETRANGELO (2024), *La dimensione plurale della cybersicurezza: da potere invisibile a processo collaborativo*, in "Rivista italiana di informatica e diritto", vol. 6, 2024, n. 2
- S. PIETROPAOLI (2024), *Un'occasione (forse) mancata. Considerazioni sulla revisione dei reati informatici proposta con il DDL Cybersicurezza*, in "Rivista italiana di informatica e diritto", vol. 6, 2024, n. 1
- M.C. POLLICINO (2025), *Gli effetti della "sommatoria" tra il GDPR e il nuovo Regolamento sulle intelligenze artificiali nell'ambito dell'attività amministrativa*, in "Rivista italiana di informatica e diritto", vol. 7, 2025, n. 1
- A. SANTOSUOSSO, G. SARTOR (2022), *La giustizia predittiva: una visione realistica*, in "Giurisprudenza italiana", 2022, n. 7
- G. SARTOR (2022-A), *Corso di informatica giuridica. L'informatica giuridica e le tecnologie dell'informazione*, Giappichelli, 2022
- G. SARTOR (2022-B), *L'intelligenza artificiale e il diritto*, Giappichelli, 2022
- A. SIMONCINI (2019), *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in "BioLaw Journal – Rivista di BioDiritto", 2019, n. 1
- G. TADDEI ELMI, A. CONTALDO (a cura di) (2024), *Intelligenza artificiale. AI Act – Regolamento (UE) 2024/1689. Il nuovo scenario giuridico europeo*, Pacini Giuridica, 2024
- G. TADDEI ELMI, S. MARCHIAFAVA (2024), *Sviluppi recenti in tema di intelligenza artificiale e diritto. Una rassegna di legislazione, giurisprudenza e dottrina (maggio-ottobre 2024)*, in "Rivista italiana di informatica e diritto", vol. 6, 2024, n. 2
- V. VASTA (2025), *Diritto dell'Unione europea e intelligenza artificiale. Riflessi sul procedimento penale*, in F. Basile, M. Biasi, L. Camaldo, G. Caneschi, B. Fragasso, D. Milani (a cura di), "Intelligenza artificiale. Diritto, giustizia, economia ed etica", Giappichelli, 2025
- G. ZICCARDI (2022), *Diritti digitali. Informatica giuridica per le nuove professioni*, Raffaello Cortina, 2022
- G. ZICCARDI, P. PERRI (a cura di) (2019), *Tecnologia e diritto, vol. III, Informatica giuridica avanzata*, Giuffrè Francis Lefebvre, 2019