

Intelligenza artificiale e governance pubblica: limiti e funzioni del diritto penale

***Artificial intelligence and public
governance: limits and functions
of Criminal Law***

Lorenza Ricci*

** Esperta in Cybersecurity*

Abstract

Technological progress often outpaces society and legal frameworks, challenging traditional approaches to governance. Artificial intelligence (Ai), as a pervasive and disruptive technology, exemplifies this tension, demanding a rethinking of legal boundaries and regulatory strategies. This article examines the role of Criminal Law in the public regulation of Ai and the management of its associated risks, analyzing its limits, potential, and interactions with other legal and regulatory tools. It highlights key features of Ai relevant to policy makers, including its dual-use nature, opacity, automation, and capacity to expand and transform threats, particularly in the context of criminal activity. The discussion addresses structural constraints of Criminal Law that hinder adaptive responses to rapidly evolving technologies, while highlighting its capacity to protect fundamental legal interests and deter harmful behavior within broader regulatory frameworks. The article also evaluates potential techniques for criminalization of malicious Ai use, while critically reviewing Italian legislative initiatives, notably Law 132/2025. It concludes that Criminal Law should play a residual, complementary role, emphasizing the need of proactive, multi-level governance, technical standards, and public awareness to promote responsible Ai use and mitigate emerging risks.

Keywords: Artificial intelligence, Criminal Law, Legal challenges, Regulation and governance, Accountability.

Introduzione

Il progresso tecnologico non bussa: irrompe, travolge e impone cambiamenti prima che la società e il diritto possano prepararsi ad accoglierli. Il diritto – cui è affidato l'oneroso compito di regolare la pacifica convivenza sociale e garantire la tutela degli interessi giuridicamente rilevanti – si mostra sempre più affannato nel tentativo di sostenere il ritmo dell'innovazione.

Tale cronico disallineamento tra la lentezza dello sviluppo normativo e la rapidità del progresso tecnologico si manifesta oggi in tutta la sua evidenza, esacerbato dall'irruzione dell'intelligenza artificiale (Ia), una tecnologia tanto dirompente e pervasiva da essere considerata rivoluzionaria. L'avvento dell'Ia è, infatti, destinato a ridefinire i confini dell'esistenza umana, insinuandosi – più o meno consapevolmente – in ogni ambito dell'esperienza quotidiana, individuale e collettiva.

In questo scenario, non bisogna arrendersi all'impetuosità del progresso tecnologico: costruire un'efficace normativa sull'Ia è fondamentale non solo per sfruttarne appieno le potenzialità, ma anche per tutelare interessi politici, strategici ed economici.

Il Legislatore e le diverse branche del diritto non possono limitarsi ad osservare, dovendo invece interrogarsi sul proprio ruolo e sulla propria capacità di interagire con una trasformazione tanto profonda quanto rapida.

La costruzione di una solida corazza normativa è essenziale per non soccombere a ciò che viene audacemente definito shock da modernità (Stella, 2003) ed impedire all'innovazione di travolgere principi, istituzioni e beni giuridici fondamentali.

Anche il diritto penale è chiamato ad intraprendere un processo critico di autovalutazione dell'adeguatezza delle proprie risorse a fronteggiare la rivoluzione tecnologica in atto. L'intrinseca imprevedibilità, l'opacità, la crescente automazione dei più sofisticati sistemi di Ia, nonché la commistione tra attività umana e attività della macchina, mettono infatti in discussione la capacità del diritto penale di intervenire efficacemente, rispettando al contempo i principi cardine del sistema.

In tale contesto, il presente contributo muove da un interrogativo: quale ruolo può svolgere il diritto penale nella regolazione dell'intelligenza artificiale e nella governance dei rischi ad essa correlati?

A tal riguardo, la tesi che si intende sostenere è che il diritto penale, per la rigidità dei suoi principi e la sua limitata capacità di adattamento a fenomeni tecnologici complessi e in rapida evoluzione, non sia idoneo a rivestire un ruolo centrale nella governance dell'intelligenza artificiale. Esso deve essere piuttosto ricondotto alla sua fisiologica funzione di extrema ratio, intervenendo solo in via residuale, all'interno di un quadro regolatorio multilivello fondato su strumenti tecnici, amministrativi e civilistici, orientati il più possibile alla prevenzione e gestione ex ante dei rischi.

In questa prospettiva, il contributo si articolerà in tre sezioni principali, volte a verificare la fondatezza della tesi proposta: la prima descriverà le caratteristiche dell'Ia rilevanti per il policy maker e i rischi connessi; la seconda analizzerà limiti e potenzialità del diritto penale come strumento di governance, esplorando le tecniche di criminalizzazione potenzialmente applicabili; la terza approfondirà il caso italiano e le strategie regolatorie attualmente adottate dal decisore politico, con un breve focus sulle sfide poste dall'uso dell'Ia da parte della pubblica amministrazione.

A valle di tale analisi, sarà possibile verificare se e in quale misura la tesi proposta trovi conferma, valutando il ruolo effettivamente esercitabile dal diritto penale nella gestione dei rischi derivanti dall'Ia, nonché l'effettiva idoneità delle misure prospettate a ridurre il divario tra l'inafferrabile progresso tecnologico e l'evoluzione normativa.

In questo senso, l'articolo intende stimolare una riflessione consapevole, senza la pretesa di fornire risposte definitive a un fenomeno in perpetua evoluzione.

1. Le peculiarità dell'Ia rilevanti per il policy maker e i rischi ad esse connessi

Il discorso sull'intelligenza artificiale è avvolto da un alone di incertezza semantica e indefinitezza concettuale – testimoniato dall'assenza stessa di una definizione condivisa del concetto di “intelligenza artificiale” – che mal si concilia con il rigore tradizionalmente associato alle discipline scientifiche¹.

Ferma restando l'assenza di una definizione universalmente riconosciuta, un punto di riferimento normativo è offerto dal Regolamento europeo sull'intelligenza artificiale (Ai Act), il cui art. 3 qualifica i sistemi di Ia come sistemi automatizzati, dotati di diversi livelli di autonomia, capaci di generare output – quali previsioni, contenuti, raccomandazioni o decisioni – idonei a influenzare ambienti fisici o virtuali².

All'interno di tale nozione è possibile ed utile distinguere, in via semplificata, tra intelligenza artificiale “tradizionale” e intelligenza artificiale generativa.

La prima, sviluppata sin dalla seconda metà del secolo scorso, si fonda su algoritmi deterministici progettati per eseguire compiti specifici in base a regole ben definite, risultando particolarmente efficace in attività di analisi dei dati, riconoscimento di pattern e decisione in contesti relativamente stabili.

L'Ia generativa rappresenta invece un'evoluzione più recente, basata su tecniche avanzate di machine learning e deep learning, capace di generare contenuti nuovi – quali testi, immagini, video e persino codici software – attraverso processi di apprendimento dai dati. A differenza dell'Ia “tradizionale”, l'Ia generativa non si limita ad applicare regole preimposte, ma sviluppa autonomamente modelli capaci di generare output originali e adattivi³.

Tale differenza incide in modo significativo sulle modalità di regolazione: mentre l'Ia tradizionale si presta più agevolmente all'inquadramento nei modelli giuridici

1. La data di nascita dell'espressione “intelligenza artificiale” non è così recente come si potrebbe ipotizzare: infatti, risale al 1955, quando John McCarthy, assistente universitario di matematica al Dartmouth College di Hanover, New Hampshire, la definì come “la scienza e l'ingegneria della creazione di macchine intelligenti e programmi per computer particolarmente intelligenti”.
2. Regolamento (Ue) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (Ce) n. 300/2008, (Ue) n. 167/2013, (Ue) n. 168/2013, (Ue) 2018/858, (Ue) 2018/1139 e (Ue) 2019/2144 e le direttive 2014/90/Ue, (Ue) 2016/797 e (Ue) 2020/1828 (Regolamento sull'intelligenza artificiale).
3. Sebbene Ia tradizionale e Ia generativa abbiano obiettivi e metodi di funzionamento diversi, entrambe sono destinate a lavorare insieme per risolvere problemi complessi e favorire l'innovazione. L'Ia tradizionale fornisce per esempio stabilità, prevedibilità e controllo, mentre la Ia generativa apre la strada alla creatività e alla generazione di nuove idee.

esistenti, l'Ia generativa accentua i problemi di opacità, autonomia e imputazione delle responsabilità.

Chiarita questa distinzione, appare dunque necessario delineare le principali caratteristiche comuni ai sistemi di Ia: ciò è essenziale al fine di comprendere a pieno le sfide – attuali e prospettiche – che questa tecnologia pone al legislatore, nonché per individuare gli strumenti più adeguati a fronteggiarle.

Una prima caratteristica dell'Ia è la sua spiccata ambivalenza: essa dischiude rilevanti benefici e nuove opportunità applicative ma, al tempo stesso, espone a rischi inediti e potenzialmente gravi, che richiedono un'attenta valutazione pubblica.

In questo senso, l'Ia può essere considerata una tecnologia “dual use”, espressione tipicamente utilizzata per indicare strumenti suscettibili di impiego sia in ambito civile sia militare. Tuttavia, nel caso dell'intelligenza artificiale, tale ambivalenza assume un significato più ampio: essa può infatti essere utilizzata tanto per finalità lecite e socialmente utili, quanto per scopi dannosi o illeciti (Brundage, Avin, Clark & Toner, 2018).

Si tratta di una caratteristica difficilmente arginabile. La maggior parte dei compiti che risulta utile automatizzare presenta una natura intrinsecamente duale⁴.

In un simile scenario, l'intenzione dell'utente finale del sistema di Ia assume un ruolo determinante: se il doppio volto dell'Ia non è eliminabile, una delle armi più potenti per contrastare l'uso malevolo dell'Ia risiede nella diffusione di una consapevolezza collettiva sulle potenzialità e i rischi connessi a questa tecnologia, volta a favorirne un utilizzo responsabile e orientato a fini leciti.

Un'ulteriore criticità dell'Ia discende dalla capacità di automatizzare e delegare attività tradizionalmente fondate sull'interazione diretta tra individui⁵: ciò consente di mantenere l'anonimato e aumenta la distanza psicologica ed emotiva tra chi utilizza il sistema di intelligenza artificiale e i soggetti su cui tale impiego impatterà. Ne consegue una ridotta percezione da parte dell'utente delle conseguenze delle proprie azioni, in quanto non direttamente coinvolto nell'interazione fisica o emotiva con l'altro soggetto. Tale effetto di desensibilizzazione, di natura prevalentemente psicologica, può favorire gli utilizzi

-
4. Basti pensare che non vi sono differenze significative tra l'uso di un drone per consegnare pacchi e lo sfruttamento dello stesso strumento per consegnare esplosivi, essendo il sistema sottostante esattamente il medesimo.
 5. Ad esempio, comunicare, osservare, prendere decisioni, o semplicemente essere presenti contestualmente con altri soggetti.

malevoli dell'Ia⁶, rafforzando l'esigenza di affiancare alle strategie di controllo e regolamentazione anche politiche di alfabetizzazione collettiva.

Una terza caratteristica critica è costituita dall'opacità dei sistemi di Ia, nota come "black-box problem". L'espressione si riferisce alla difficoltà di ricostruire e spiegare il processo attraverso cui un sistema giunge a una determinata decisione, ossia il passaggio dall'input all'output.

Il tema assume particolare rilievo sotto il profilo dell'affidabilità: se non è possibile comprendere come il sistema perviene ad una determinata conclusione, è naturale interrogarsi sul grado di fiducia che possa essergli accordato⁷ (Von Eschenbach, 2021).

Nella prospettiva del superamento dell'opacità si colloca lo sviluppo della cosiddetta "explainable Ai" (Xai), ossia di modelli progettati per rendere i sistemi intelligenti più trasparenti, interpretabili e spiegabili⁸. Tuttavia, neppure i modelli di Xai sono in grado di assicurare la piena trasparenza: le spiegazioni fornite possono risultare complesse e difficilmente accessibili ai soggetti privi di competenze tecniche, restringendo la comprensibilità effettiva dei sistemi a una platea limitata di esperti. L'obiettivo finale è quello di sviluppare sistemi di Ia interamente comprensibili alla totalità dei soggetti coinvolti nel loro impiego.

Le caratteristiche brevemente illustrate non costituiscono meri aspetti tecnici, ma incidono in modo significativo sulla natura e sull'intensità dei rischi connessi agli impieghi illeciti dell'Ia, imponendo una valutazione attenta delle minacce emergenti e delle strategie idonee a governarle.

Tali rischi possono essere efficacemente rappresentati attraverso la metafora dell'iceberg: se allo stato attuale risultano visibili soltanto alcune manifestazioni di uso distorto dell'Ia, esiste una parte sommersa – ben più ampia – destinata ad affiorare progressivamente allo sviluppo e alla diffusione capillare dei sistemi intelligenti (Miró-Llinares, Duvac, Toader & Galarza, 2024).

Una prima implicazione correlata all'impiego malevolo dell'intelligenza artificiale è rappresentata dalla capacità di espandere e amplificare le minacce esistenti. Osservata

-
6. Un esempio emblematico è rappresentato dai sistemi d'arma autonomi: un criminale avrà certamente meno remore morali nel commettere un omicidio utilizzando simili strumenti che, a differenza di una qualsiasi arma tradizionale, consentono di superare lo scoglio psicologico di essere fisicamente presente sulla scena del crimine e dover affrontare direttamente la vittima.
 7. La problematica si accentua considerando la crescente tendenza a delegare decisioni di capitale rilevanza all'Ia, anche in settori sensibili quali la salute, la sicurezza e la salvaguardia di beni umani rilevanti.
 8. Alcuni modelli mirano a rivelare il funzionamento interno della "scatola nera", mentre altri si limitano a rendere comprensibile la decisione finale a chi osserva dall'esterno, senza fornire una ricostruzione completa dei processi decisionali sottostanti.

attraverso la lente penalistica, l'Ia favorisce l'introduzione di veri e propri "gemelli digitali" di reati tradizionali, senza necessariamente introdurre condotte ontologicamente nuove, ma incidendo in modo significativo sulle modalità di commissione degli illeciti (*Ibidem*).

Più nello specifico, lo sviluppo dell'intelligenza artificiale contribuisce all'espansione delle minacce in un triplice senso: amplia la platea di soggetti in grado di condurre attacchi o illeciti in generale; agevola l'aumento del tasso di successo delle operazioni illecite; estende i potenziali bersagli⁹.

In sostanza, l'intelligenza artificiale contribuisce ad una vera e propria "democratizzazione" della capacità di commettere reati, consentendo a chiunque di compiere illeciti che in precedenza erano prerogativa di soggetti dotati di elevate conoscenze e competenze specialistiche¹⁰. All'interno di questo quadro gioca un ruolo significativo la capacità dell'intelligenza artificiale di generare una sensibilizzazione emotiva nell'agente, riducendo l'empatia verso la vittima e, parallelamente, accrescendo la predisposizione a commettere l'illecito¹¹.

Accanto all'espansione delle minacce già note, l'uso illecito dell'intelligenza artificiale è destinato a favorire l'emersione di nuove tipologie di rischio e, in particolare, nuove tipologie di attacchi, che non sarebbe possibile realizzare con le sole capacità umane¹². Inoltre, ulte-

-
9. Un esempio emblematico di minaccia in espansione è costituito dagli attacchi di *spear phishing*. Letteralmente "pesca con la lancia" è una truffa tramite cui i cybercriminali spingono le vittime a rivelare informazioni sensibili, come le proprie credenziali di accesso. In alcuni casi, i truffatori riescono ad ottenere l'accesso ai dati tramite link o allegati malevoli, che, una volta aperti, installano malware nel computer della vittima. Ciò che contraddistingue lo *spear phishing* dal più generico phishing è la natura mirata di questo tipo di attacco. Gli attacchi *spear phishing* vengono perpetrati tramite l'invio di messaggi altamente personalizzati, che rendono le e-mail credibili agli occhi delle vittime ed aumentano la probabilità che l'utente apra link o allegati malevoli. L'uso dell'Ia consente oggi di automatizzare l'intero processo di identificazione delle potenziali vittime, ricerca delle reti sociali e professionali e creazione di messaggi plausibili e pertinenti, abbattendole barriere tecniche e permettendo a un numero crescente di individui di realizzare attacchi sofisticati.
 10. Un esempio lampante di tale democratizzazione è rappresentato dalla creazione di deep fake pornografici, la quale non richiede alcuna competenza specialistica, grazie alla disponibilità di numerose applicazioni che consentono a qualunque utente di generare falsi contenuti pornografici a partire da un'immagine. È stato infatti rilevato che sarebbero sufficienti appena venticinque minuti e nessun dispendio economico per produrre un video deep fake porno della durata di un minuto raffigurante una persona qualsiasi, partendo semplicemente da un'immagine nitida del volto della vittima.
 11. Ad esempio, è possibile prevedere un aumento degli attacchi tramite droni o sistemi d'arma autonomi in generale, con un conseguente ampliamento delle minacce associate a tali strumenti.
 12. Per citare un esempio, l'essere umano – per sua natura – non è generalmente in grado di imitare realisticamente la voce altrui o creare manualmente contenuti audio che somiglino in tutto e per tutto a registrazioni di discorsi umani. L'intelligenza artificiale consente la creazione di sistemi di sintesi vocale

riori minacce possono derivare dalla presenza di vulnerabilità dei sistemi di Ia non ancora risolte. Queste debolezze – intrinseche ai sistemi in fase di sviluppo o non adeguatamente protetti – potrebbero essere sfruttate per colpire gli utenti che ne fanno uso¹³.

Da ultimo, l'impiego dell'intelligenza artificiale è destinato ad alterare profondamente il carattere tipico delle minacce oggi conosciute. I progressi in ambito dell'Ia sono infatti idonei ad accrescere l'efficacia degli attacchi, rendendoli sempre più precisi e difficili da attribuire ad un soggetto determinato (Brundage, Avin, Clark & Toner, 2018). Suddetta alterazione può essere esaminata attraverso tre principali linee direttrici.

In primo luogo, è prevedibile che gli attacchi altamente efficaci diventeranno più tipici e frequenti. L'uso dell'Ia consentirà di superare il tradizionale compromesso tra frequenza e scala degli attacchi, rendendo possibile la realizzazione di operazioni simultaneamente più incisive, ripetute e su larga scala¹⁴.

In secondo luogo, grazie alle proprietà dell'efficienza e scalabilità¹⁵, l'Ia consentirà di raggiungere un livello di precisione senza precedenti nell'identificazione dei potenziali bersagli, contribuendo alla diffusione di attacchi sempre più selettivi¹⁶.

capaci di imitare la voce umana con un livello di realismo tale da renderle indistinguibili da quelle autentiche in assenza di misure di autenticazione appositamente progettate. Tali strumenti aprono scenari inediti per la diffusione di disinformazione, con potenziali ripercussioni gravi sull'ordine pubblico, sui processi democratici e sulla fiducia collettiva.

13. Un esempio di sfruttamento delle vulnerabilità dei sistemi di intelligenza artificiale riguarda i veicoli a guida autonoma: un criminale potrebbe indurre il sistema ad interpretare erroneamente determinati segnali, provocando un incidente. Ad esempio, potrebbe indurre l'auto a non riconoscere un semaforo rosso, impedendole di attivare il freno. Oppure ancora, un'immagine di un cartello di stop con alcune modifiche minime nei pixel, che gli esseri umani continuerebbero a riconoscere come un segnale di stop, potrebbe essere interpretata in modo completamente sbagliato da un sistema di Ia.
14. Ad esempio, lo *spear phishing* – un tipo di attacco già menzionato precedentemente – è certamente più efficace rispetto al generico phishing, il quale non prevede la personalizzazione dei messaggi, ma è al contempo più costoso e non può essere condotto su larga scala. Al contrario, gli attacchi di phishing generici, nonostante gli inferiori tassi di successo, sono redditizi perché sferrati su larga scala.
15. Un sistema di intelligenza artificiale è efficiente se, una volta addestrato, è in grado di portare a termine un determinato compito in modo più rapido e/o con minori costi rispetto all'ipotesi in cui lo stesso compito sia svolto da un essere umano. Invece, un sistema di Ia è scalabile se, dato un determinato compito, aumentando la potenza di calcolo o aumentando le copie del sistema stesso, sarà in grado di portare a termine molte più istanze del compito assegnatogli. Per fare un esempio, un sistema di riconoscimento facciale è sia efficiente che scalabile in quanto, una volta sviluppato e addestrato, potrà essere applicato a molte telecamere diverse e ad un costo di gran lunga inferiore rispetto a quello che richiederebbe l'assunzione di analisti umani chiamati a svolgere un compito equivalente.
16. Ad esempio, sarà possibile utilizzare sciami di droni che utilizzano la tecnologia del riconoscimento facciale al fine di colpire specifici individui, piuttosto che utilizzare forme di violenza generiche e indiscriminate.

Un terzo profilo è rappresentato dalla capacità dell'intelligenza artificiale di agevolare l'anonimato dei potenziali attori. La questione si è posta anche prima dell'avvento dell'Ia in relazione ai reati commessi nel cyberspazio, ove la garanzia dell'anonimato costituisce da sempre una delle principali problematiche, alimentando la percezione di impunità degli attaccanti nonché la convinzione di non essere individuati. In tale contesto, l'Ia è destinata a complicare i meccanismi di attribuzione della responsabilità, aggravando una criticità già ampiamente sentita dalla comunità degli esperti.

Le considerazioni sin qui svolte evidenziano la marcata complessità del tema, soprattutto alla luce del ruolo sempre più pervasivo che l'intelligenza artificiale è destinata ad assumere nella vita quotidiana degli individui. In tale prospettiva, appare necessario superare una concezione meramente tecnologica dell'Ia, riconoscendone piuttosto la natura di vera e propria infrastruttura organizzativa, idonea a incidere in modo significativo sulla configurazione e sull'evoluzione dei processi decisionali, tanto in ambito pubblico quanto in quello privato.

La letteratura sui sistemi informativi ha evidenziato come l'impiego di sistemi algoritmici favorisca l'emersione di forme di algorithmic management, nelle quali decisioni tradizionalmente rimesse all'autorità umana vengono progressivamente delegate a modelli automatizzati, incidendo sui meccanismi di esercizio del potere e sui processi di legittimazione delle decisioni. Ne consegue che i sistemi di intelligenza artificiale devono ormai essere intesi come "istituzionalmente incorporati" (*institutionally embedded*) all'interno di strutture complesse, nelle quali l'output finale è il risultato dell'interazione tra una pluralità di attori, tra cui sviluppatori, fornitori di dati, integratori e utenti finali. (Pasquale, 2015; Kitchin, 2017).

In questo quadro emergono rilevanti criticità, tra le quali, come si vedrà, assume particolare rilievo quella relativa all'attribuzione della responsabilità. Essa, infatti, non risulta più agevolmente riconducibile a un unico centro decisionale, ma richiede l'elaborazione di modelli interpretativi di tipo distribuito e multilivello, capaci di cogliere la dimensione socio-tecnica che caratterizza i sistemi di intelligenza artificiale.

2. Limiti e potenzialità del diritto penale come strumento di governance

Al fine di stabilire se, in quale misura, e con quali modalità, il diritto penale sarà eventualmente chiamato ad intervenire nella governance dell'intelligenza artificiale e dei suoi impieghi illeciti, è necessario individuarne preliminarmente i principali limiti e potenzialità.

L'impiego del diritto penale come strumento di governo dei rischi connessi all'Ia incontra limiti strutturali che non dipendono da carenze contingenti del sistema, bensì

dalla sua stessa architettura assiologica. I principi cardine della materia penale, pur costituendo presidi irrinunciabili di garanzia, ne riducono significativamente la capacità di adattamento a fenomeni tecnologici caratterizzati da rapidità evolutiva, opacità e imprevedibilità.

In primo luogo, il principio di tassatività, insieme al divieto di analogia in *malam partem* che ne discende, si pone in evidente tensione con l'intelligenza artificiale, i cui impieghi illeciti assumono forme mutevoli e spesso difficilmente riconducibili a schemi tipizzati ex ante¹⁷.

Ne deriva un duplice rischio: da un lato, l'inadeguatezza delle fattispecie esistenti a ricomprendere nuove modalità di offesa senza forzature interpretative; dall'altro, la tentazione di introdurre norme eccessivamente ampie o indeterminate, con conseguente compromissione delle garanzie di precisione e certezza del diritto.

In questo senso, la tassatività si configura come un limite invalicabile all'uso "elastico" del diritto penale per fronteggiare l'innovazione tecnologica.

Un secondo profilo critico riguarda il principio di colpevolezza. L'attribuzione della responsabilità penale presuppone la possibilità di muovere un rimprovero personale per una condotta lesiva di un bene giuridico: requisito difficilmente conciliabile con l'impiego di sistemi di intelligenza artificiale, caratterizzati dalla capacità di automazione, auto-apprendimento nonché opacità dei processi interni. L'impossibilità di riconoscere una soggettività penalmente rilevante in capo ai sistemi di Ia e, al tempo stesso, la difficoltà di individuare un unico soggetto umano responsabile, rendono il paradigma penalistico tradizionale solo parzialmente idoneo a intercettare le dinamiche dell'illecito tecnologico.

Le criticità appena evidenziate emergono con particolare chiarezza nel passaggio dal piano teorico a quello applicativo. A titolo esemplificativo, si consideri l'impiego di sistemi di intelligenza artificiale nel triage sanitario, finalizzati a supportare il personale medico nella prioritarizzazione dei pazienti sulla base della gravità delle condizioni cliniche. In tali contesti, l'adozione di modelli algoritmici consente di incrementare l'efficienza e la rapidità delle decisioni, ma espone al contempo al rischio di errori sistemici, bias nei dati e opacità nei criteri di classificazione. Qualora un sistema di Ia attribuisca erroneamente una priorità inferiore a un paziente in condizioni critiche, determinando un ritardo nell'intervento e un aggravamento delle condizioni di salute, si pongono rilevanti questioni in ordine all'attribuzione della responsabilità.

17. Il principio di tassatività rappresenta un'espressione rafforzata del principio di legalità. Letto in correlazione con il principio di riserva di legge, il principio di tassatività sancisce l'impossibilità di applicare le norme incriminatrici al di fuori dei casi espressamente previsti dalla legge.

In una simile ipotesi, l'evento lesivo non sarebbe agevolmente riconducibile alla condotta di un singolo soggetto, ma appare piuttosto riconnesso al funzionamento complessivo del sistema, quale esito dell'interazione tra sviluppatori, fornitori di dati, struttura sanitaria e personale medico. Ciò evidenzia l'inadeguatezza dei modelli penalistici tradizionali, fondati su un rapporto diretto tra autore e fatto, e rafforza la necessità di un approccio multilivello (v. *infra*).

Infine, il principio di sussidiarietà – che impone al diritto penale di operare in extrema ratio¹⁸ – evidenzia forse il limite più rilevante in una prospettiva di governance pubblica. La regolazione dell'intelligenza artificiale si presta, per sua natura, a strumenti preventivi, flessibili e adattivi, quali standard tecnici, obblighi organizzativi, controlli amministrativi e meccanismi di accountability ex ante. In tale contesto, il ricorso alla sanzione penale si giustifica soltanto dinanzi ad offese particolarmente gravi e non adeguatamente fronteggiabili con misure meno afflittive. Attribuire al diritto penale un ruolo centrale nella disciplina dell'Ia rischierebbe di tradirne la funzione residuale e sfociare in interventi inefficaci.

In sintesi, il principale limite del diritto penale nella governance dell'Ia risiede proprio nella rigidità dei suoi principi fondativi: principi che devono essere preservati in quanto presidi di garanzia, ma che al contempo impediscono di concepire la risposta penalistica come soluzione primaria o autosufficiente rispetto ai rischi posti dall'uso illecito dei sistemi intelligenti.

Ciononostante, il diritto penale conserva alcune innegabili potenzialità che ne giustificano un impiego selettivo nella regolazione pubblica dell'Ia. In particolare, esso è in grado di offrire una tutela rafforzata ai beni giuridici fondamentali – quali la vita, l'integrità fisica, la libertà personale e la sicurezza pubblica – che possono essere significativamente incisi dall'uso illecito o distorto di sistemi intelligenti, intervenendo a presidio di soglie di rischio oltre le quali la risposta repressiva appare necessaria. A ciò si affianca la funzione di prevenzione generale della sanzione penale, idonea, se inserita in un quadro normativo coerente e prevedibile, ad orientare ex ante i comportamenti degli attori coinvolti nello sviluppo e nell'impiego dell'Ia, incentivando pratiche responsabili e una maggiore attenzione ai profili di rischio.

Alla luce di tali considerazioni, occorre sviluppare in via preliminare un quadro normativo integrato – di natura extra-penale, amministrativa e civile – idoneo a disciplinare gli aspetti più tecnici della materia. Solo successivamente sarà possibile identificare le condotte potenzialmente meritevoli di sanzione penale, così da relegare l'intervento repressivo alla sua funzione residuale di strumento di chiusura del sistema regolatorio.

18. Cioè, soltanto quando non siano disponibili altre misure meno afflittive in grado di garantire l'adeguata tutela di determinati interessi.

3. Limiti e potenzialità del diritto penale come strumento di governance

La necessità di relegare l'intervento penalistico ad una funzione residuale nella governance dell'Ia non implica l'esclusione aprioristica dell'intervento penale, bensì la sua collocazione in una fase successiva ed eventuale, laddove gli strumenti settoriali – predisposti ed attuati – risultino insufficienti a garantire una tutela effettiva ed adeguata degli interessi esposti ai rischi della rivoluzione tecnologica in corso.

Laddove si confermasse la legittimità e l'esigenza di un intervento penalistico, è possibile ipotizzare alcune tecniche di criminalizzazione delle condotte illecite connesse all'intelligenza artificiale. Ciò nella piena consapevolezza che un intervento efficace non può fondarsi su soluzioni generalizzate, come la creazione di fattispecie incriminatrici applicabili indiscriminatamente a tutti i sistemi di Ia, ma richiede una valutazione puntuale delle specificità tecniche e del contesto applicativo di ciascun sistema, al fine di assicurare risposte normative mirate e proporzionate alla natura del rischio.

In tale prospettiva, l'attenzione ricade inevitabilmente sul già richiamato Ai Act. Quest'ultimo adotta uno specifico approccio, basato appunto sul rischio (cosiddetto risk-based approach), con una forte vocazione per il principio di precauzione. Il Regolamento distingue, infatti, tra sistemi a rischio inaccettabile – espressamente vietati (art. 5) – sistemi ad alto rischio, sottoposti a stringenti obblighi di conformità, e sistemi a rischio limitato o minimo, per i quali sono previsti requisiti più attenuati.

In particolare, per i sistemi ad alto rischio, l'Ai Act impone una serie di obblighi tecnici e organizzativi in capo agli operatori della filiera – tra cui produttori, fornitori e utilizzatori – quali l'adozione di sistemi di gestione del rischio, requisiti di qualità e governance dei dati, documentazione tecnica, tracciabilità delle operazioni, nonché meccanismi di supervisione umana.

Sulla base di questa premessa e mantenendo come riferimento la cornice normativa europea, appare in primo luogo ipotizzabile l'introduzione di reati contro la creazione e l'uso di sistemi di intelligenza artificiale vietati.

Sebbene l'Ai Act non imponga agli Stati membri di intervenire penalmente al fine di criminalizzare le pratiche di Ia vietate, è inevitabile la sua forte influenza nelle future scelte di criminalizzazione.

Invero, le categorie di rischio introdotte dall'Ai Act si rivelano fondamentali al fine di individuare l'ambito di legittimo intervento del diritto penale. In questa prospettiva, la regolamentazione extra-penale – in particolare quella tecnico-amministrativa – è chiamata a svolgere una funzione primaria nella definizione delle soglie del cosiddetto

“rischio accettabile” e nella predisposizione dei presidi di prevenzione. Solo all’interno di una simile cornice regolatoria sarà possibile collocare in modo coerente e sistematicamente giustificato l’intervento penalistico, quale strumento residuale destinato a operare nei casi di violazione significativa degli obblighi imposti o di esposizione a rischi non tollerabili per beni giuridici fondamentali (Miró-Llinares, Duvac, Toader & Galarza, 2024).

La seconda tecnica di criminalizzazione muove da una presa di coscienza già evidenziata in precedenza: la complessità delle catene decisionali, la molteplicità dei soggetti coinvolti, la crescente automazione, imprevedibilità e opacità dei sistemi di Ia, determinano un progressivo allontanamento tra la condotta umana e l’evento lesivo, rendendo difficoltoso sia l’accertamento del nesso di causalità, sia l’individuazione di un unico soggetto responsabile.

Al fine di arginare tali problematiche, è possibile ipotizzare un’anticipazione della tutela penale, introducendo reati di pericolo – e, dunque, di mera condotta – volti a sanzionare la mera realizzazione od omissione di determinati comportamenti connessi alla progettazione, produzione e impiego di sistemi di Ia idonei a cagionare un danno a beni giuridici rilevanti, indipendentemente dall’effettiva verifica di un evento lesivo. In questo modo verrebbe risolto il problema relativo al nesso di causalità, che non necessita di essere accertato in relazione ai reati di mera condotta¹⁹.

La difficoltà di identificare un unico soggetto responsabile potrebbe, invece, essere affrontata imputando la responsabilità direttamente all’organizzazione che ha prodotto o utilizzato il sistema di Ia pericoloso. Ciò imporrebbe una revisione a più livelli della disciplina di cui d.lgs. 231/2001²⁰, finalizzata, tra le altre cose, ad ampliare il novero dei reati presupposto²¹, nonché, e soprattutto, a svincolare la responsabilità dell’ente dall’esistenza di una persona fisica autrice del reato, così da estendere l’operatività della normativa anche agli eventi lesivi riconducibili all’azione automatica e talvolta imprevedibile dei sistemi di intelligenza artificiale.

19. Tale anticipazione deve in ogni caso essere ancorata alla regolamentazione extra-penale, chiamata a definire gli standard tecnici e di sicurezza da rispettare nello sviluppo dei sistemi di intelligenza artificiale, nonché le condizioni di rischio da prevenire, i livelli di rischio consentito, le misure cautelari da adottare al fine di circoscriverli. La disciplina contenuta in queste normative costituirà lo strumento principale per stabilire con chiarezza il momento a partire dal quale l’anticipazione dell’intervento penalistico potrebbe risultare legittima, nonché in relazione a quali interessi giuridici. Anche in questo contesto, il principale riferimento normativo per il legislatore nazionale sarà rappresentato dall’Ai Act.

20. Decreto legislativo 8 giugno 2001, n. 231, Disciplina della responsabilità amministrativa delle persone giuridiche, delle società e delle associazioni anche prive di personalità giuridica, a norma dell’art. 11 della legge 29 settembre 2000, n. 300.

21. Includendovi nuove fattispecie penalmente rilevanti connesse all’impiego dell’intelligenza artificiale.

Nondimeno, una simile impostazione, sebbene idonea a risolvere in astratto talune criticità imputative, rischierebbe di produrre effetti distortivi sul piano economico e dell'innovazione, disincentivando lo sviluppo e l'adozione di tecnologie basate sull'intelligenza artificiale per il timore di un'eccessiva esposizione al rischio penale.

Al fine di mitigare le criticità evidenziate, appare dunque opportuno adottare un modello di responsabilità distribuita nell'impiego dei sistemi di intelligenza artificiale, fondato sull'attribuzione a ciascun attore coinvolto di ruoli e obblighi chiaramente definiti. Una simile impostazione consente di rafforzare i meccanismi di prevenzione e di controllo ex ante, riducendo al massimo la necessità di ricorrere all'intervento del diritto penale.

La Tav. 1 schematizza questa ideale distribuzione, evidenziando, per ciascuna categoria di attori, i principali compiti, gli obblighi minimi e i rimedi amministrativi e civilistici esperibili in via preventiva rispetto all'attivazione della risposta penale.

Tav. 1 (fonte: elaborazione degli autori).

Soggetto	Responsabilità principale	Obblighi minimi	Rimedi prima dell'intervento penale
Sviluppatore	Progettazione e correttezza dell'algoritmo	Documentazione tecnica, validazione del modello, registro delle modifiche	Correzione del software, revisione tecnica, aggiornamento dei protocolli
Fornitore dati	Qualità, accuratezza e integrità dei dati	Registrazione fonti dati, verifiche di coerenza, protezione della privacy	Rettifica dati errati, obblighi di comunicazione, accordi contrattuali
Integratore	Inserimento dell'Ia nei sistemi operativi o processi decisionali	Test di integrazione, monitoraggio, reportistica sull'implementazione	Adeguamento dei processi, programmi di formazione e capacity building, audit interno
Rup / Dirigente	Supervisione e approvazione dei sistemi Ia	Verifica della conformità normativa, governance del progetto, documentazione decisionale	Modifica procedure, sanzioni interne, richieste di aggiornamento del sistema
Operatore/utilizzatore	Applicazione corretta delle procedure e uso dell'Ia	Formazione specifica, registro delle decisioni, obblighi di segnalazione e meccanismi di escalation	Correzione operativa, feedback, meccanismi di contestazione

Tornando alle possibili tecniche di criminalizzazione dell'Ia, la terza opzione si configura come una specificazione della precedente, fondata sulla difficoltà di provare la volontarietà delle condotte di produttori, distributori o utilizzatori di Ia e propone di introdurre reati di pericolo colposi – che nell'ordinamento italiano rappresentato un *numerus clausus* – che sanzionino la violazione involontaria degli standard di diligenza richiesti dalla normativa di settore²².

L'ultima tecnica di criminalizzazione, la più semplice ed intuitiva, consiste nell'introduzione di nuove circostanze aggravanti da applicarsi alle ipotesi in cui il reato venga commesso con l'impiego dei sistemi di Ia, valorizzando la capacità di tali strumenti di amplificare l'offensività delle condotte illecite.

Per quanto intuitiva, questa tecnica non deve rappresentare l'approccio generale per contrastare il maggior danno derivante dall'utilizzo dell'intelligenza artificiale. Considerata la diffusione ormai capillare dell'utilizzo della tecnologia nella commissione dei reati, un ricorso generalizzato a nuove aggravanti rischierebbe infatti di condurre ad un'applicazione sistematica della forma aggravata, riducendo significativamente lo spazio per una valutazione caso per caso (Gullo & Flor, 2024). Ne deriva l'esigenza di limitare l'introduzione di circostanze aggravanti a ipotesi specifiche, valutando puntualmente l'effettiva incidenza dell'Ia sulla gravità del singolo reato.

Le criticità sin qui evidenziate emergono con particolare chiarezza quando si passa dal piano teorico a quello applicativo.

4. Le soluzioni proposte dal Legislatore italiano: analisi critica della legge 132/2025

Alla luce dell'analisi svolta, è possibile offrire una sintetica panoramica delle principali direttrici di intervento e di adattamento dell'ordinamento italiano, non esclusivamente penale, che appaiono maggiormente significative rispetto alle criticità emerse.

Il punto di riferimento è senza dubbio rappresentato dalla legge 132/2025, recante “Disposizioni e deleghe al Governo in materia di intelligenza artificiale”, che costituisce

22. A ben vedere, numerosi sono gli scenari concreti in cui si potrebbe rendere necessaria l'introduzione di una responsabilità penale per colpa. Un esempio significativo è offerto dai veicoli a guida autonoma e dagli incidenti da essi causati. In effetti, tale problema era già stato ravvisato dall'Assemblea parlamentare del Consiglio d'Europa, che aveva sollevato il problema dell'attribuzione della responsabilità, non essendo possibile imputare il fatto né all'essere umano a bordo del veicolo, né al produttore, considerando che il veicolo autonomo potrebbe causare un incidente pur operando in conformità con il progetto e le norme applicabili.

il primo quadro normativo nazionale volto a disciplinare in modo organico una materia tanto complessa e in costante evoluzione qual è l'Ia.

Nata come disegno di legge 1146/2924, presentato nel maggio 2024, la legge è stata definitivamente approvata il 17 settembre 2025 ed entrata in vigore il successivo 10 ottobre. Tuttavia, come suggerito dal titolo stesso, la legge si caratterizza per un ampio ricorso alla delega governativa, che richiederà di attendere altri dodici mesi per l'adozione dei relativi decreti legislativi attuativi.

La precisazione sull'iter di approvazione della legge e sui prossimi sviluppi appare doverosa poiché riporta in primo piano il tema del cronico ritardo dell'evoluzione normativa rispetto all'inafferrabile progresso tecnologico. Nell'attesa che il Governo adempia alla delega, l'imprevedibile e irrefrenabile evoluzione dell'intelligenza artificiale espone al rischio evidente che, una volta approvate tutte le modifiche, queste risultino già obsolete.

Fatta questa premessa, alla legge 132/2025 va senz'altro riconosciuto più di qualche merito. Questa, infatti rappresenta la prima iniziativa legislativa finalizzata ad allineare l'ordinamento interno di uno Stato membro alle disposizioni introdotte dall'AI Act.

Si tratta senza dubbio di una chiara scelta politica e strategica, volta a dimostrare la consapevolezza da parte dello Stato italiano sia della rilevanza dell'Ia come risorsa destinata a ridefinire gli assetti socioeconomici, sia della necessità di fissare regole precise e capaci di orientare e trainare lo sviluppo e l'utilizzo di questa rivoluzionaria tecnologia.

Venendo brevemente all'analisi critica delle principali direttrici della legge in esame, l'articolo di apertura esplicita immediatamente il duplice approccio che connota la normativa: da un lato, promuovere un impiego "corretto, trasparente e responsabile" dell'intelligenza artificiale – sempre in una dimensione antropocentrica – al fine di coglierne a pieno le potenzialità; dall'altro lato, garantire un adeguato presidio rispetto ai rischi "economici e sociali" connessi all'utilizzo dell'Ia nonché al relativo impatto sui diritti fondamentali²³.

23. Tale approccio viene ribadito e specificato nell'art. 3, che enuncia i principi generali che devono orientare l'intero ciclo di vita dei sistemi d'intelligenza artificiale. In particolare, la norma stabilisce che, dalla fase di ricerca e sperimentazione, fino all'effettivo utilizzo dei sistemi e modelli di Ia, è necessario garantire il rispetto dei diritti fondamentali e delle libertà riconosciute dall'ordinamento italiano ed europeo, oltre che i principi di trasparenza, proporzionalità, sicurezza, protezione dei dati personali, riservatezza, robustezza, accuratezza, non discriminazione e sostenibilità (commi 1 e 2). Inoltre, lo stesso articolo sancisce – rafforzando quanto già enunciato dall'art.1 – che lo sviluppo e la concreta applicazione dei sistemi di intelligenza devono avvenire nel rispetto dell'autonomia e del potere decisionale dell'uomo, in piena conformità con l'approccio antropocentrico che caratterizza tanto il disegno di legge quanto il Regolamento europeo (comma 3).

Già nella relazione introduttiva all'iniziale disegno di legge emergeva chiaramente l'intento di bilanciare opportunità e rischi insiti nell'impiego dell'Ia, attraverso la previsione di principi generali e disposizioni settoriali volti, da un lato, a promuovere l'innovazione tecnologica per migliorare le condizioni di vita dei cittadini e rafforzare la coesione sociale e, dall'altro, introdurre strumenti di gestione del rischio ispirati a un approccio antropocentrico²⁴.

Sintomatico della consapevolezza dell'importanza strategica ed economica dell'Ia è anche l'art. 5, rubricato, "Principi in materia di sviluppo economico", che sancisce quali principi generali la concorrenza, l'innovazione e l'equità nel mercato dell'Ia, prevedendo inoltre che lo Stato agevoli l'accesso a dati di qualità per imprese e comunità scientifiche.

Di rilievo sono anche gli artt. 7-18, che disciplinano l'impiego dell'Ia in settori critici per lo sviluppo nazionale – come sanità, lavoro, giustizia, professioni intellettuali e pubblica amministrazione – affinché esso avvenga nel rispetto di criteri in grado di garantire la gestione e la soluzione dei rischi legati allo svolgimento di tali attività strategiche.

Con specifico riferimento alla pubblica amministrazione (Pa), la legge 132/2025 segna un cambio di paradigma, imponendo un ripensamento di strategie, procedure e competenze. In questo senso, l'alternativa per le Pa non è più impiegare o meno l'intelligenza artificiale, bensì governarla in modo consapevole, strutturato e conforme alla legge, oppure subirla in modo disorganico, rischioso e, in ultima analisi, illegittimo (Panucci, 2025).

Il Legislatore mira, infatti, a promuovere un uso dell'Ia nella pubblica amministrazione idoneo a rendere più effettivi i principi costituzionali di buon andamento, imparzialità ed efficienza (art. 97 Cost.). In questo contesto si colloca l'art. 14 della legge 132/2025, che delinea un perimetro di azione chiaro e introduce principi destinati ad incidere in modo significativo sull'approccio delle amministrazioni all'innovazione digitale.

In particolare, l'utilizzo dell'Ia da parte delle Pa dovrà perseguire finalità concrete e misurabili:

- incrementare l'efficienza della propria attività;

24. In tale prospettiva, la legge 132/2025 intende porsi in funzione complementare con l'Ai Act, intervenendo nelle sfere di competenza del diritto interno. Analogamente a quanto previsto dal Regolamento europeo, la proposta di legge in esame si prefigge di tutelare i diritti fondamentali, la democrazia, lo stato di diritto e la sostenibilità ambientale – modulando gli interventi in ragione dei rischi e dell'impatto potenziale dei sistemi di intelligenza artificiale – e, contestualmente, di promuovere l'innovazione tecnologica quale strumento di benessere per i cittadini

- ridurre i tempi di definizione dei procedimenti;
- aumentare la qualità e la quantità dei servizi erogati.

Tali obiettivi risultano indissolubilmente connessi all'obbligo di garantire la conoscibilità del funzionamento dei sistemi di Ia e la tracciabilità del loro utilizzo. Ciò conferma la piena consapevolezza del Legislatore dell'assoluta necessità di superare l'opacità che caratterizza i sistemi di Ia, esigenza che si fa particolarmente stringente nel contesto della pubblica amministrazione, la cui attività è strutturalmente informata al principio di trasparenza^{25 26}.

L'art. 14 individua, poi, un limite invalicabile nell'impiego dell'Ia da parte della pubblica amministrazione, stabilendo che tale tecnologia debba rimanere uno strumento di supporto dell'attività amministrativa, senza poter mai sostituire l'autonomia decisionale del funzionario pubblico né esonerarlo dalla sua piena responsabilità per i procedimenti e provvedimenti in cui l'Ia sia stata impiegata.

Si tratta di una disposizione di grande rilievo: il principio della responsabilità personale del funzionario²⁷, cardine del nostro diritto amministrativo, non solo viene preservato, ma risulta rafforzato, imponendo un livello ancora più elevato di consapevolezza nell'uso degli strumenti tecnologici.

Consapevole del fatto che i principi di strumentalità e responsabilità non possono reggersi da soli, il Legislatore ha previsto, al terzo comma dell'art. 14, l'obbligo per le pubbliche amministrazioni di adottare misure tecniche, organizzative e formative finalizzate a

25. La Pa non può e non potrà mai trasformarsi in una "scatola nera" algoritmica. Ogniqualvolta un sistema di Ia viene impiegato nell'esercizio dell'azione amministrativa, deve essere garantita la massima trasparenza sul suo funzionamento nonché la possibilità di ricostruire ex post il percorso logico che ha condotto a un determinato risultato.

26. Anche il d.lgs. 36/2023 (Codice dei Contratti pubblici) all'art. 30, "Uso di procedure automatizzate nel ciclo di vita dei contratti pubblici", ribadisce l'importanza della comprensibilità e tracciabilità del funzionamento dei sistemi di Ia. Tra le disposizioni di maggior rilievo si evidenzia in primo luogo, l'obbligo di garantire, nelle fasi di acquisizione e sviluppo del sistema, la disponibilità del codice sorgente e di ogni altro elemento idoneo a consentire la comprensione delle logiche sottese al funzionamento algoritmico. In secondo luogo, deve essere assicurato il rispetto dei principi di non discriminazione algoritmica, di conoscibilità e di comprensibilità, nonché il fondamentale principio della non esclusività della decisione algoritmica, il quale impone che, nell'ambito del processo decisionale, sia comunque previsto un intervento umano capace di controllare, validare o, se del caso, disattendere l'esito generato dal sistema automatizzato.

27. L'art. 28 Cost. dispone che "I funzionari e i dipendenti dello Stato e degli enti pubblici sono direttamente responsabili, secondo le leggi penali, civili e amministrative, degli atti compiuti in violazione di diritti".

garantire un utilizzo responsabile dell'intelligenza artificiale e a sviluppare le competenze trasversali degli utilizzatori.

Tale previsione testimonia la presa d'atto della necessità di intervenire prioritariamente per sviluppare una maggiore conoscenza e consapevolezza da parte degli operatori circa le potenzialità e i rischi connessi all'impiego dell'Ia. Attenzione ulteriormente ribadita nella sezione dedicata alle (numerose) deleghe al Governo, ove spicca l'invito a predisporre percorsi di alfabetizzazione e formazione per l'utilizzo dell'Ia, sia in generale sia per professionisti e studenti.

Il Legislatore riconosce altresì che lo sfruttamento dell'intelligenza artificiale da parte della pubblica amministrazione debba avvenire in un contesto di cooperazione e scambio reciproco tra soggetti pubblici e privati. In tale direzione si pone l'art. 19 della legge 132/2025 che, nell'istituire la Strategia nazionale per l'intelligenza artificiale, individua tra le relative funzioni quella di favorire la collaborazione tra le amministrazioni pubbliche e i soggetti privati nello sviluppo e adozione di sistemi di Ia, nonché di coordinare l'attività della Pa e, nuovamente, promuovere la ricerca e diffusione della conoscenza in materia.

Per quanto complessivamente coerente con il quadro normativo internazionale, la normativa nazionale presenta un limite di fondo nell'approccio stesso adottato dal legislatore: l'Ia viene ancora concepita in chiave meramente strumentale e, soprattutto, nel pieno controllo dell'essere umano creatore ed utilizzatore. Questo orientamento rischia di rivelarsi già obsoleto, considerate le capacità dei sistemi più avanzati di compiere azioni e prendere decisioni in modo automatico, talvolta imprevisto e imprevedibile da parte degli stessi sviluppatori.

Preoccupa anche la previsione secondo cui le pubbliche amministrazioni dovranno adempiere agli obblighi introdotti dalla legge "con le risorse umane, strumentali e finanziarie disponibili a legislazione vigente". Una simile clausola desta non poche perplessità: come può un processo di trasformazione tanto profondo e complesso, quale quello richiesto dall'integrazione responsabile dell'intelligenza artificiale nell'azione amministrativa, realizzarsi efficacemente in assenza di adeguati investimenti, tanto sul piano economico quanto su quello delle risorse umane e delle competenze?

Fermo restando quanto sostenuto nei paragrafi che precedono, il Legislatore ha ritenuto opportuno intervenire anche sul versante penalistico, delegando innanzitutto il Governo ad operare in un triplice ambito:

- previsione di strumenti – e correlate sanzioni – volti ad inibire la diffusione nonché a rimuovere contenuti creati illecitamente, anche con l'uso dell'Ia;

- introduzione di autonome fattispecie di reato, dolose o colpose, finalizzate a punire l'omessa adozione o adeguamento di misure di sicurezza in fase di produzione, diffusione e impiego professionale dei sistemi di intelligenza artificiale, qualora tali omissioni generino un pericolo concreto di lesione della vita, nonché dell'incolumità pubblica o individuale e della sicurezza nazionale;
- precisazione dei criteri di imputazione della responsabilità penale delle persone fisiche e degli enti in relazione agli illeciti connessi all'uso dell'intelligenza artificiale²⁸.

Accanto alla delega governativa, l'art. 26 della legge 132/2025 introduce alcune modifiche al sistema penale immediatamente operative.

In primo luogo, viene introdotta una nuova circostanza aggravante comune (art. 61 n. 11 – undecies c.p.), applicabile qualora il reato sia commesso con sistemi di Ia particolarmente insidiosi, che ostacolano la difesa pubblica o privata, o aggravano le conseguenze del fatto. Sono inoltre contemplate aggravanti speciali per specifiche fattispecie di reato, ove il fatto sia commesso con l'uso dell'Ia²⁹.

Particolarmente significativa è poi l'introduzione di un nuovo delitto, l'art. 612-quater c.p., rubricato "Illecita diffusione di contenuti generati o alterati con sistemi di intelligenza artificiale"³⁰. La disposizione si colloca nel solco della crescente attenzione e preoccupazione del legislatore in relazione al delicato fenomeno dei deep fake, che rappresenta senza dubbio uno degli utilizzi più insidiosi dell'Ia³¹.

L'elevatissimo potenziale lesivo dell'uso distorto dei deep fake, capace di compromettere beni individuali e collettivi di primaria importanza, quali la dignità personale, l'onore, la libertà sessuale, l'identità personale, la privacy, la corretta informazione e – nei casi

28. Il tutto tenendo conto effettivo grado di controllo che l'agente è in grado di esercitare sui sistemi di Ia, a dimostrazione della consapevolezza, da parte del legislatore, dell'esistenza di un tasso di imprevedibilità intrinseco ai più evoluti modelli di intelligenza artificiale.

29. Ad esempio, l'art. 294 c.p., "Attentati contro i diritti politici del cittadino" sancisce oggi al comma 2, "La pena è della reclusione da due a sei anni se l'inganno è posto in essere mediante l'impiego di sistemi di intelligenza artificiale".

30. In particolare, la norma punisce con la reclusione da uno a cinque anni "Chiunque cagiona un danno ingiusto ad una persona, cedendo, pubblicando o altrimenti diffondendo, senza il suo consenso, immagini, video o voci falsificati o alterati mediante l'impiego di sistemi di intelligenza artificiale e idonei a indurre in inganno sulla loro genuinità".

31. Si tratta di un neologismo, coniato nel 2017, che incrocia la locuzione deep learning (insieme di tecniche che permettono all'intelligenza artificiale di imparare a riconoscere le forme) con la parola "fake" (falso) e descrive quelle tecniche che permettono di combinare e sovrapporre immagini tra loro utilizzando reti neurali generative (in gergo tecnico, Gan).

più gravi – lo stesso bene vita, rendeva auspicabile l'introduzione di una nuova fattispecie incriminatrice, pur perfettabile³².

Per quanto rilevanti, neppure gli interventi di natura penalistica risultano pienamente risolutivi. Come già sottolineato, accanto all'anticipazione della tutela penale, sarebbe stato altresì necessario un immediato ripensamento del tradizionale paradigma di imputazione della responsabilità, in particolare attraverso un adeguamento della disciplina ex d.lgs. 231/2001 volto a svincolare la responsabilità della persona giuridica dall'esistenza di una persona fisica autrice del reato: su questo fronte occorrerà attendere i decreti attuativi.

In sintesi, la legge 132/2025, pur rappresentando un primo ammirevole tentativo regolatorio e pur offrendo alcuni spunti senz'altro rilevanti, non è idonea ad offrire risposte pienamente adeguate ed efficaci alle sfide poste dall'Ia. Invero – tra le altre criticità rilevate – l'eccessivo ricorso alla delega governativa per l'attuazione di alcuni delicati interventi riformatori, unitamente alla lentezza fisiologica dei processi legislativi, rischia non solo aggravare ulteriormente il ritardo dell'evoluzione normativa al cospetto dell'inafferrabile progresso tecnologico, ma anche di produrre un quadro regolatorio già obsoleto al momento della sua effettiva entrata in vigore.

5. Focus: la responsabilità delle pubbliche amministrazioni nell'impiego dei sistemi di intelligenza artificiale

Tutte le criticità analizzate nel presente contributo si manifestano in tutta la loro evidenza anche nell'impiego dell'intelligenza artificiale da parte della pubblica amministrazione. In tale ambito, i principi sanciti dall'art. 14 della legge 132/2025³³, nonché dall'art. 30 del Codice dei Contratti pubblici, pur costituendo presidi fondamentali a garanzia di un uso dell'Ia conforme ai principi cardine dell'azione amministrativa, incontrano rilevanti difficoltà applicative derivanti, in particolare da un duplice ordine di fattori: da un

32. Il tema è particolarmente delicato, considerato che il 98% dei video deep fake è rappresentato da materiale a contenuto pornografico, ottenuto tramite l'incorporazione estremamente realistica del volto di un soggetto su corpi nudi o in contesti di natura sessualmente esplicita. Secondo un report del gruppo informatico Home Security House, tra il 2022 e il 2023 la produzione di deep fake pornografici è aumentata del 464%, evidenziando una crescita esponenziale del fenomeno. Si registra inoltre un incremento dello scarto tra deep fake pornografici e generici, in quanto i primi costituiscono il 98% del totale (Fonte: <https://keepnetlabs.com/blog/deepfake-statistics-and-trends>, ultimo aggiornamento, marzo 2026). Parimenti preoccupanti sono i dati pubblicati da "Deep Strike", che segnala come il numero di deep fake diffusi sia passato dai 500 mila del 2023 agli 8 milioni del 2025, corrispondente ad un tasso di crescita del 900% l'anno. (Fonte, <https://deepstrike.io/blog/deepfake-statistics-2025>, settembre 2025 e <https://keepnetlabs.com/blog/deepfake-statistics-and-trends>, ultimo aggiornamento, marzo 2026)

33. Nonché dall'art. 30 del Codice dei Contratti pubblici.

lato, la forzatura di concepire l'Ia in chiave meramente strumentale, nel pieno controllo dell'essere umano che la impiega; dall'altro, la difficoltà di individuare un unico centro di imputazione della responsabilità per eventuali danni cagionati da un sistema di intelligenza artificiale.

Come già ampiamente sottolineato, anche quando è una Pa ad impiegare un sistema di intelligenza artificiale il risultato finale si configura come il prodotto dell'interazione di una pluralità di soggetti, sia pubblici sia privati: ideatori dell'algoritmo, sviluppatori del software, fornitori dei dati, funzionari che si avvalgono del sistema e, infine, l'amministrazione utilizzatrice.

Nell'attesa di un intervento legislativo volto a chiarire in modo definitivo la ripartizione delle responsabilità, appare tuttavia necessario individuare un criterio di imputazione che assicuri una tutela quanto più possibile effettiva dei destinatari dell'azione amministrativa.

In questa prospettiva, alla luce del principio di immedesimazione organica, è possibile affermare che la Pa che si avvale dell'Ia sarà responsabile del danno patito dal soggetto sul quale si producono gli effetti negativi del provvedimento algoritmico, ferma restando la possibilità di rivalersi sui singoli responsabili, per quanto difficilmente individuabili.

Ma non è tutto: ferma restando la necessità di relegare l'intervento penalistico ad una funzione residuale, resta complessa l'individuazione del regime di responsabilità applicabile. Poiché il Legislatore continua a qualificare l'Ia come strumento di supporto all'attività provvedimentale, non sembrano emergere, allo stato, significative deroghe rispetto al modello ordinario della responsabilità da provvedimento illegittimo, tradizionalmente ricondotto alla responsabilità extracontrattuale, salvo specifiche ipotesi – come nel settore degli appalti pubblici – in cui l'ordinamento ammette forme di responsabilità oggettiva.

Accanto alla responsabilità verso l'esterno può infine configurarsi una responsabilità amministrativa da danno erariale connessa all'impiego di sistemi di intelligenza artificiale. Anche tale forma di responsabilità presenta, tuttavia, rilevanti difficoltà applicative, sia per la frequente impossibilità di individuare con certezza la componente del sistema che ha causato il danno e, di conseguenza, quale tra i soggetti coinvolti ne sia effettivamente responsabile; sia per la complessità di accertare l'esistenza dell'elemento soggettivo del dolo o della colpa grave, richiesto per la sua configurabilità, in un contesto caratterizzato da un elevato livello di imprevedibilità.

Nel complesso, le problematiche esaminate confermano come l'impiego dell'intelligenza artificiale da parte della pubblica amministrazione riproponga, in termini accentuati, le criticità strutturali della governance dell'Ia, evidenziando la necessità di un intervento normativo effettivamente risolutivo.

Riflessioni conclusive

Le considerazioni svolte nel presente contributo, volto ad indagare il ruolo che il diritto penale può effettivamente svolgere nella regolazione pubblica dell'intelligenza artificiale e nella governance dei rischi ad essa connessi, conducono ad una conclusione netta, in linea con la tesi sostenuta sin dall'inizio: il diritto penale, con i suoi rigidi principi e la sua scarsa capacità reattiva dinanzi alle nuove esigenze imposte dall'evoluzione tecnologica, non appare idoneo a rivestire un ruolo centrale nella gestione delle problematiche legate all'avvento dell'intelligenza artificiale.

I principi di tassatività, di precisione, di colpevolezza – sacrosanti pilastri della materia penalistica – si pongono in potenziale frizione con l'indeterminatezza, opacità e imprevedibilità che permeano una tecnologia in inesorabile evoluzione qual è l'Ia. È tuttavia soprattutto in forza del principio di sussidiarietà che si impone la necessità di relegare l'intervento penale ad una funzione residuale.

In tale prospettiva, ogni eventuale riforma del catalogo di fattispecie incriminatrici vigenti dovrebbe essere preceduta dalla predisposizione di un solido quadro tecnico-normativo integrato, di natura extra-penale, volto a definire in modo puntuale gli standard tecnici da rispettare nello sviluppo dei sistemi di intelligenza artificiale, le condizioni di rischio da prevenire, le misure di mitigazione da adottare nonché le soglie del cosiddetto “rischio accettabile”, anche alla luce dell'evoluzione della normativa europea.

Complementare a tale processo sarà l'attuazione della Strategia italiana per l'intelligenza artificiale, orientata alla promozione di uno sviluppo dell'Ia sicuro, etico e inclusivo.

Piuttosto che affannarsi in una perenne rincorsa agli sviluppi dell'Ia e alla repressione dei relativi usi illeciti, appare dunque necessario adottare un approccio proattivo e preventivo, fondato su una cooperazione multidisciplinare e multilivello tra regole giuridiche e misure tecniche. Una simile impostazione consentirebbe un migliore coordinamento tra i diversi settori dell'ordinamento, favorendo la creazione di un quadro tecnico-normativo improntato alla massima chiarezza, coerenza ed efficacia.

Ciò non implica, come già sottolineato, l'esclusione aprioristica dell'intervento penale, bensì la sua collocazione in una fase successiva ed eventuale, laddove gli strumenti settoriali – predisposti ed attuati – risultino insufficienti a garantire una tutela effettiva ed adeguata degli interessi giuridici coinvolti.

In definitiva, il contrasto all'uso illecito dell'intelligenza artificiale richiede un impegno condiviso da parte di tutte le branche del diritto e della tecnica, nonché in generale, di tutti gli attori della società contemporanea: decisori politici, esperti e ricercatori, imprese e, in ultima istanza, l'intera collettività.

In questa prospettiva, la chiave di volta nella prevenzione e contrasto all'uso illecito dell'Ia risiederà proprio nei consociati e in un'adeguata alfabetizzazione collettiva capace di promuovere una consapevolezza diffusa e costantemente aggiornata circa le potenzialità e i rischi insiti nell'impiego dei sistemi intelligenti, inesorabilmente destinati a penetrare in ogni aspetto dell'esperienza quotidiana.

In un contesto segnato da un'inevitabile asimmetria tra sviluppo tecnologico e capacità regolatoria, la consapevolezza umana resta la risorsa più preziosa per fronteggiare le derive più insidiose della rivoluzione tecnologica.

Riferimenti Bibliografici

- Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (Xai). *IEEE*, 6.
- Azzali, V., & Ellecosta, N. (2023). La questione deepfake in Italia: una panoramica. *Media Laws*, 3.
- Brundage, M., Avin, S., Clark, J., & Toner, H. (2018). The malicious use of artificial intelligence: Forecasting prevention and mitigation. *arXiv*, 2.
- Ciancaglini, V. (2020). Malicious uses and abuses of artificial intelligence, *Trend Micro Research; United Nations Interregional Crime and Justice Research Institute (Unicri); Europol's European Cybercrime Centre (Ec3)*, 11.
- Cimini, S. (2026). La responsabilità civile e amministrativa di fronte alle sfide dell'intelligenza artificiale. *Ceridap*, 1
- Consulich, F. (2024). Il diritto penale al tempo dell'intelligenza artificiale. Prospettive punitive nazionali dopo l'AI Act. *Diritto di difesa*, 12.
- Cummings, M. L. (2004). Creating Moral Buffers in Weapon Control Interface Design. *IEEE Technology and Society Magazine*, 23.
- Deepstrike (2025). Deepfake Statistics 2025: AI Fraud Data & Trends. Consultabile in <https://deepstrike.io/blog/deepfake-statistics-2025>
- De Flammineis, S. (2024). Fattispecie penali nel contesto dell'intelligenza artificiale. Lo spunto del d.d.l. 1146/2024. *Sistema penale*, 9.
- Dwivedi, R., et. al. (2023). Explainable Ai (Xai): Core Ideas, Techniques, and Solutions. *ACM Computing Surveys*, 55(9).
- Fiandaca, G., & Musco, E. (2019). *Diritto penale parte generale* (ottava edizione). Bologna: Zanichelli.
- Fragasso, B. (2023). La responsabilità penale del produttore di sistemi di intelligenza artificiale. *Diritto penale contemporaneo – Riv. Trim.*, 1.
- Gullo, A., & Flor, R. (2024). Italian report on criminalisation of ai-related offences, in Miró-Llinares, F., Duvac, C., Toader, T., & Galarza M.S. (Eds.), Criminalisation of AI-related offences. *RIDP*, 95 (1).
- Herley, C. (2010). The Plight of the Targeted Attacker in a World of Scale. Workshop on the Economics of Information Security.
- Italiano, G. F. (2019). Intelligenza artificiale, che errore lasciarla agli informatici. agendadigitale.eu
- Keepnet (2026). Deepfake Statistics & Trends 2026: Growth, Risks, and Future Insights. Consultabile in <https://keepnetlabs.com/blog/deepfake-statistics-and-trends>
- Kitchin, R. (2016). Thinking critically about and researching algorithms. *Information, communication & society*, 20.
- Marinucci, G., Dolcini, E., & Gatta, G. L. (2022). *Manuale di diritto penale parte generale* (undicesima edizione). Milano: Giuffrè.
- McCarthy, J., Minsky, M. K., Rochester, N., & Shannon, C. E. (1955). A proposal for the Dartmouth summer research project on artificial intelligence.
- Miró-Llinares, F., Duvac, C., Toader, T., & Galarza, M. S. (Eds.) (2024). Criminalisation of AI-related offences. *RIDP*, 95(1).
- Ordonselli, N. (2021). “Porn Deepfake”: profili di diritto penale. *Cyberlaws*, 1.

- Panucci, G. (2025). Legge 132/2025 sull'intelligenza artificiale: un punto di svolta per la Pa italiana. Consultabile in <https://www.ilpersonale.it/legge-132-2025-sullintelligenza-artificiale-un-punto-di-svolta-per-la-pa-italiana/>
- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, Massachusetts: Harvard University Press.
- Picotti, L., & Panattoni, B. (Eds.) (2023). Categorie tradizionali del diritto penale e intelligenza artificiale: crisi o palingenesi? *RIDP*, 94.
- Raita, Y., Goto, T., Faridi, M. K., Brown, D. F. M., Camargo Jr., C. A., & Hasegawa, K. (2019). Emergency department triage prediction of clinical outcomes using machine learning models. *Critical Care*, 23.
- Senato della Repubblica (2024). Relazione introduttiva al disegno di legge 1146/2024. Consultabile in <https://www.senato.it/service/PDF/PDFServer/BGT/01418921.pdf>
- Stella, F. (2003). *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*. Milano: Giuffrè.
- Von Eschenbach, W. J. (2021). Transparency and the Black Box Problem: Why We Do Not Trust AI. *Philosophy and Technology*, 34.
- Wörner, L. (2024). German national report on criminalisation of AI-related offences, in Miró-Llinares, F., Duvac, C., Toader, T., & Galarza, M. S. (Eds.) (2024). Criminalisation of AI-related offences. *RIDP*, 95(1).