Editoriale Scientifica

OSSERVATORIO POLITICHE PUBBLICHE PER LE AUTONOMIE

Anticipazione
Fascicolo 1/2026

www.osservatorioentilocali.unirc.it

Intelligenza artificiale e responsabilità penale: causalità, colpevolezza e tutela anticipata del rischio tecnologico. Dalle categorie tradizionali alle nuove prospettive normative

di Andrea Battaglia

Osservatorio Politiche pubbliche per le Autonomie

Tutti i contributi, ad eccezione dell'editoriale di apertura, sono stati sottoposti a referaggio (*blind peer review*)
.

Direttore**:** prof. Francesco Manganaro, professore ordinario di Diritto amministrativo e pubblico, Università Mediterranea di Reggio Calabria

Condirettori
Prof. Giuseppe Tropea, prof.ssa Vincenza Caracciolo La Grotteria

Comitato scientifico
Francesco Astone, Jean Bernard Auby, Estefania Maria de Quiroz Barboza; Antonio Barone, Vittoria Berlingò, Carlos Botassi, Giuseppe Caia, Marco Calabrò, Marco Cammelli, Vincenzo Caputi Jambrenghi, Guido Corso, Fulvio Cortese, Angela Costaldello, Carlos Balbín, Carlos Botassi, Maria Teresa Carballeira Rivera, Sabino Cassese, Mario P. Chiti, Maria Cristina Cavallaro, Giovanna Colombini, Giandomenico Comporti, Alfredo Contieri, Maria De Benedetto, Giacinto della Cananea, Daria de Pretis, Marina D'Orsogna, Marco Dugato, Enrico Follieri, Fabrizio Fracchia, Fabio Francario, Gabriella De Giorgi Cezzi, Diana Urania Galetta, Carlo Emanuele Gallo, Loredana Giani, Marcella Gola, Guido Greco, Nicola Gullo, Maria Immordino, Roberta Lombardi, Marco Mazzamuto, Fabio Merusi, Giuseppe Morbidelli, Remo Morzenti Pellegrini, Nino Paolantonio, Sergio Perongini, Giuseppe Piperata, Aristide Police, Pier Luigi Portaluri, Margherita Ramajoli, Mauro Renna, Anna Romeo, Fabio Saitta, Aldo Sandulli, Maria Alessandra Sandulli, Franco Gaetano Scoca, Anna Simonati, Mario Spasiano, Fabrizio Tigano, Michele Trimarchi, Bruno Tonoletti, Francesco Tuccari, Jacques Ziller.

Comitato editoriale
Rocco Parisi; Giuseppe Carlo; Cristina Fragomeni; Francesca Saraceno; Debora Bellocco

Editoriale Scientifica

Intelligenza artificiale e responsabilità penale: causalità, colpevolezza e tutela anticipata del rischio tecnologico. Dalle categorie tradizionali alle nuove prospettive normative

Andrea Battaglia

Abstract

La crescente diffusione dei sistemi di intelligenza artificiale caratterizzati da autonomia decisionale e opacità algoritmica pone interrogativi di particolare complessità per il diritto penale, soprattutto con riferimento all'accertamento del nesso causale e alla formulazione del giudizio di colpevolezza. Il contributo esamina tali questioni muovendo da una delimitazione funzionale dell'oggetto di indagine, circoscritto ai sistemi di intelligenza artificiale autonomi, interattivi e caratterizzati da un funzionamento opaco, riconducibile al paradigma della *black box*. L'analisi si concentra sulla possibilità di valorizzare il criterio della probabilità logica e di reinterpretare il rapporto tra condotta umana ed evento attraverso la categoria della causalità psichica, senza compromettere le garanzie proprie dell'imputazione penale. Sul versante della colpevolezza, viene approfondito il ruolo del rischio consentito, delle regole cautelari e degli obblighi di controllo umano, con particolare riguardo al Regolamento (UE) 2024/1689. Il lavoro affronta, inoltre, la crisi del diritto penale d'evento e la prospettiva di fattispecie di tutela anticipata, nonché le potenzialità del d.lgs. 8 giugno 2001, n. 231, e del suo art. 8 per l'imputazione della responsabilità agli enti. In conclusione, le innovazioni introdotte dal d.lgs. 9 settembre 2026, n. 160, sono considerate alla luce delle questioni teoriche esaminate, allo scopo di verificare in quale misura il nuovo quadro normativo confermi l'esigenza di un modello di responsabilità fondato sulla gestione del rischio tecnologico e sulla permanenza della centralità dell'agente umano.

Parole chiave: intelligenza artificiale; diritto penale; *black box*; causalità psichica; colpevolezza; rischio consentito; tutela anticipata; responsabilità degli enti; d.lgs. n. 231/2001.

1. Premessa metodologica: una nozione di intelligenza artificiale penalisticamente orientata

L'integrazione dei sistemi di intelligenza artificiale nei settori maggiormente sensibili della vita sociale impone di verificare la tenuta delle categorie tradizionali del diritto penale. L'impiego di sistemi capaci di apprendere dai dati, adattarsi all'ambiente e produrre risultati non integralmente predeterminati dal programmatore può incidere sulle modalità

attraverso le quali vengono realizzate condotte penalmente rilevanti e, soprattutto, sulle possibilità di attribuire a un soggetto umano un evento lesivo.

La questione non consiste, tuttavia, nel domandarsi in termini generali se l'intelligenza artificiale sia compatibile con il diritto penale. Una simile impostazione rischierebbe di sovrapporre fenomeni tecnologici profondamente differenti, accomunati dalla sola denominazione di intelligenza artificiale. Un algoritmo deterministico, un sistema di classificazione e un modello di apprendimento automatico capace di generare risposte variabili presentano, infatti, caratteristiche che non possono essere considerate equivalenti ai fini dell'imputazione della responsabilità.

È pertanto necessario delimitare l'oggetto dell'indagine attraverso una nozione funzionale e penalisticamente orientata di intelligenza artificiale. Il presente contributo circoscrive l'analisi ai sistemi autonomi, interattivi e caratterizzati da opacità decisionale, nei quali l'*output* non costituisce la semplice esecuzione di una sequenza di istruzioni interamente predeterminata e verificabile dall'essere umano.

L'autonomia rilevante in questa prospettiva non deve essere confusa con una presunta volontà della macchina, né implica il riconoscimento di una soggettività giuridica autonoma. Essa individua, piuttosto, uno scarto tra le condizioni iniziali predisposte dall'uomo e il risultato concretamente prodotto dal sistema. L'interattività, a sua volta, indica la capacità del sistema di ricevere informazioni dall'ambiente, elaborarle e produrre risposte che incidono sul contesto esterno o sui processi decisionali umani.

Il tratto maggiormente problematico è rappresentato dalla cosiddetta *black box*: l'opacità del processo attraverso il quale il sistema giunge a un determinato risultato. Tale opacità può rendere difficoltosa la ricostruzione *ex post* della sequenza decisionale, anche quando siano conosciuti l'architettura generale del sistema, i dati utilizzati e gli obiettivi perseguiti.

La delimitazione proposta consente di evitare due opposti errori metodologici. Il primo consiste nell'estendere indiscriminatamente le problematiche dell'autonomia algoritmica a qualsiasi applicazione informatica. Il secondo consiste nel considerare l'opacità tecnologica una ragione sufficiente per abbandonare le categorie fondamentali del diritto penale. L'obiettivo deve essere, al contrario, quello di individuare i limiti applicativi dei modelli tradizionali e le condizioni entro le quali essi possano continuare a operare.

Da questa impostazione discende una prima conclusione: l'autonomia algoritmica non determina automaticamente l'irresponsabilità dell'essere umano, ma impedisce di fondare la responsabilità sulla sola circostanza che un soggetto abbia sviluppato, immesso sul mercato o utilizzato un sistema successivamente coinvolto nella produzione di un evento lesivo.

2. Il problema causale: probabilità logica e causalità psichica

Il principale ostacolo alla costruzione di un modello di responsabilità penale per i danni riconducibili all'intelligenza artificiale è rappresentato dall'accertamento del nesso causale. Nei sistemi di apprendimento automatico, infatti, la relazione tra *input* e *output* può non essere ricostruibile attraverso una legge scientifica capace di descrivere in termini deterministici il comportamento della macchina.

Il problema assume particolare rilievo quando l'evento sia il risultato dell'interazione tra più fattori: progettazione del sistema, selezione e qualità dei dati, fase di addestramento, aggiornamenti successivi, modalità di impiego e condizioni concrete dell'ambiente operativo. In tali circostanze, la pluralità dei soggetti coinvolti e la distribuzione delle rispettive competenze possono rendere particolarmente complessa l'individuazione del contributo causale individuale.

Non sarebbe corretto, tuttavia, concludere che l'incertezza del funzionamento algoritmico determini necessariamente l'impossibilità di accertare il nesso causale. Occorre distinguere l'assenza di una spiegazione completa del processo interno alla macchina dall'impossibilità di dimostrare, secondo i criteri propri del giudizio penale, la rilevanza causale di una determinata condotta umana.

In questa prospettiva, assume rilievo il criterio della probabilità logica, elaborato dalla giurisprudenza penale in particolare a partire dalla sentenza delle Sezioni unite della Corte di cassazione nel caso Franzese. L'accertamento causale non può essere ridotto alla verifica di una correlazione statistica, ma richiede una valutazione individualizzata delle circostanze del caso concreto, delle spiegazioni alternative e della capacità della ricostruzione proposta di escludere ragionevoli ipotesi concorrenti.

Applicato ai sistemi di intelligenza artificiale, tale criterio consente di mantenere ferma la distinzione tra probabilità statistica e certezza processuale, evitando che l'opacità tecnologica si traduca tanto in un'indiscriminata attribuzione di responsabilità quanto in un'automatica esclusione della stessa.

La difficoltà emerge soprattutto quando il comportamento della macchina non sia integralmente prevedibile neppure da parte del programmatore. In simili ipotesi, la condotta umana può non costituire la causa deterministica dell'evento, ma può avere contribuito a orientare il processo decisionale del sistema, predisponendone le condizioni operative, gli obiettivi e i parametri di funzionamento.

Da qui la proposta di valorizzare, sul piano teorico, la categoria della causalità psichica, adattandone criticamente le potenzialità al rapporto tra *input* umano e *output* algoritmico. L'idea non è quella di assimilare la macchina a una mente umana, né di trasferire automaticamente al sistema categorie elaborate per spiegare il comportamento delle persone. Si tratta, piuttosto, di interrogarsi sulla possibilità di descrivere una relazione nella quale l'intervento umano orienta un processo decisionale autonomo senza determinarne necessariamente, in termini meccanici, ogni singolo passaggio.

La causalità psichica, così intesa, può costituire una chiave interpretativa per comprendere come una decisione umana iniziale continui a essere rilevante nella genesi dell'evento, nonostante la successiva elaborazione autonoma delle informazioni da parte del sistema.

Una simile prospettiva incontra, tuttavia, un limite invalicabile: non può sostituire la dimostrazione del nesso causale richiesta dal diritto penale. L'orientamento del processo algoritmico non equivale, di per sé, alla prova che la condotta contestata abbia causato l'evento. Occorre sempre verificare, attraverso un accertamento rigoroso e individualizzato, quale contributo abbia effettivamente fornito l'agente e se tale contributo possa essere collegato all'evento secondo i criteri giuridici applicabili.

La causalità psichica deve dunque essere intesa come uno strumento di approfondimento della relazione tra condotta umana e processo algoritmico, non come una scorciatoia argomentativa capace di colmare le lacune della prova. La complessità tecnologica impone di affinare gli strumenti dell'accertamento, non di abbassarne gli standard.

3. Colpevolezza e rischio consentito: la responsabilità come gestione del rischio tecnologico

Le difficoltà causali si accompagnano a un problema altrettanto rilevante sul versante della colpevolezza. Anche quando sia possibile dimostrare il contributo causale di una condotta umana, resta da accertare se l'evento sia personalmente rimproverabile all'agente.

Il rischio è quello di costruire forme di responsabilità fondate sulla posizione occupata all'interno della filiera tecnologica. Il produttore potrebbe essere chiamato a rispondere perché ha progettato il sistema; il programmatore perché ne ha sviluppato il codice; l'utilizzatore perché ne ha consentito l'impiego. Una simile impostazione finirebbe per attribuire rilievo decisivo alla mera partecipazione al processo tecnologico, prescindendo dalla concreta violazione di un dovere cautelare e dalla possibilità di prevedere ed evitare l'evento.

Per scongiurare tale deriva, il giudizio di colpevolezza deve essere ricostruito intorno alla gestione del rischio tecnologico. La questione non è se un sistema di intelligenza artificiale possa produrre risultati dannosi, circostanza connaturata all'impiego di tecnologie complesse, ma se il rischio concretizzatosi fosse riconoscibile e se l'agente disponesse di strumenti adeguati per prevenirlo o contenerlo.

In questo quadro assume particolare rilievo la categoria del rischio consentito. Lo sviluppo e l'impiego di tecnologie innovative comportano inevitabilmente margini di rischio che non possono essere eliminati integralmente senza rendere impossibile qualsiasi attività tecnologica. Il diritto penale non può trasformarsi in uno strumento di garanzia assoluta contro ogni evento avverso, né può pretendere che il produttore assicura un risultato integralmente privo di rischi.

4. La crisi del diritto penale d'evento e la prospettiva della tutela anticipata

La difficoltà di ricostruire il nesso causale e di accertare la colpevolezza non costituisce soltanto un problema applicativo, destinato a emergere nella soluzione dei singoli casi concreti. Essa rivela una tensione più profonda tra le caratteristiche della società tecnologica e la struttura tradizionale del diritto penale d'evento. Quest'ultimo, pur nelle differenti articolazioni assunte dalle singole fattispecie incriminatrici, si fonda su una sequenza ricostruttiva nella quale una condotta umana, attiva od omissiva, viene posta in relazione con un evento naturalistico o giuridicamente rilevante, al fine di stabilire se quest'ultimo possa essere attribuito all'agente sul piano oggettivo e, successivamente, se sussistano i presupposti per muovergli un rimprovero sul piano soggettivo. Nei reati colposi d'evento, tale accertamento richiede, inoltre, di verificare la violazione di una regola cautelare, la prevedibilità e l'evitabilità dell'esito lesivo, nonché la realizzazione nel caso concreto del rischio che la regola violata mirava a prevenire.

Questa architettura risponde a una precisa esigenza di garanzia: impedire che la mera successione temporale tra una condotta e un risultato dannoso, oppure la semplice constatazione che un soggetto occupasse una determinata posizione organizzativa, siano sufficienti a fondare una responsabilità penale. L'evento non può essere imputato per il solo fatto di essersi verificato dopo una determinata condotta; occorre dimostrare, secondo i criteri propri dell'imputazione oggettiva e soggettiva, che quel risultato sia riconducibile all'agente. Il paradigma del diritto penale d'evento, pertanto, non rappresenta un ostacolo da rimuovere, bensì un insieme di vincoli che delimitano legittimamente il potere punitivo. La questione che si pone nell'ambito dell'intelligenza artificiale non consiste, dunque, nel superamento di tali garanzie, ma nella verifica della loro effettiva capacità di operare quando la produzione dell'evento è mediata da sistemi tecnici complessi, capaci di elaborare informazioni, adattare le proprie prestazioni e generare risultati non immediatamente ricostruibili attraverso una relazione lineare tra l'azione umana iniziale e l'esito finale.

La trasformazione tecnologica incide, anzitutto, sulla rappresentazione del rapporto tra condotta ed evento. Nei modelli più semplici di interazione uomo-macchina, l'esito del processo può essere ricondotto, almeno in linea di principio, a una sequenza sufficientemente intelligibile di istruzioni, operazioni e risultati. Nei sistemi di intelligenza artificiale autonomi, interattivi e caratterizzati da opacità algoritmica, invece, la decisione o il comportamento che precedono immediatamente l'evento possono derivare dall'elaborazione di una pluralità di dati, parametri e correlazioni, il cui contributo specifico non è sempre agevolmente isolabile. L'*output* non coincide necessariamente con la mera esecuzione di un comando umano puntualmente determinato: può rappresentare il risultato di un processo computazionale nel quale l'intervento iniziale dell'uomo, la configurazione del sistema, i dati impiegati, gli aggiornamenti successivi e le interazioni con l'ambiente concorrono secondo modalità non sempre integralmente prevedibili.

In questo scenario, la difficoltà non è soltanto individuare la causa materiale dell'evento, ma comprendere quale significato giuridico attribuire ai diversi contributi che hanno concorso

alla sua produzione. Il progettista può aver definito l'architettura del sistema; il produttore può aver scelto determinate soluzioni tecniche; il soggetto che lo impiega può averlo destinato a un contesto operativo specifico; l'organizzazione può aver stabilito procedure di controllo insufficienti; l'utilizzatore professionale può aver omesso un intervento di supervisione. Nessuno di questi contributi, considerato isolatamente, conduce necessariamente all'evento lesivo. Al contempo, l'eventuale assenza di una singola condotta immediatamente riconoscibile come causa dell'evento non autorizza a concludere che quest'ultimo sia privo di una spiegazione causalmente rilevante o che debba rimanere estraneo a qualsiasi forma di responsabilità. Impone, piuttosto, di ricostruire con rigore il contributo effettivamente prestato da ciascun soggetto e il ruolo svolto dal sistema nel processo che ha condotto al risultato[1].

È in questo passaggio che emerge una prima tensione con il diritto penale d'evento. Il modello tradizionale richiede di muovere dall'evento verificatosi per risalire alla condotta umana causalmente rilevante, accertando se, eliminando mentalmente quest'ultima, il risultato sarebbe venuto meno e se la conclusione sia sostenuta da un adeguato fondamento scientifico. Nei processi mediati dall'intelligenza artificiale, tuttavia, la ricostruzione controfattuale può incontrare difficoltà peculiari: l'eliminazione di una determinata condotta umana non sempre consente di ricostruire in modo attendibile quale *output* il sistema avrebbe prodotto in un contesto altrimenti identico; inoltre, il comportamento del sistema può dipendere dall'interazione di variabili numerose, da condizioni ambientali mutevoli e da processi di elaborazione non completamente trasparenti. Ciò non determina l'inapplicabilità dei criteri causali tradizionali, né legittima la loro sostituzione con mere correlazioni statistiche. Evidenzia, piuttosto, la necessità di adattare l'indagine probatoria alle peculiarità del fenomeno tecnologico, senza rinunciare alla verifica del nesso causale individuale e al livello di rigore richiesto in sede penale[2].

La medesima tensione investe la colpevolezza. La responsabilità per un evento lesivo presuppone che esso sia personalmente rimproverabile all'agente, sulla base di una condotta che questi avrebbe potuto e dovuto tenere diversamente. Nei contesti tecnologici complessi, tuttavia, la distanza tra il momento della decisione umana e quello della manifestazione dell'evento può ampliarsi considerevolmente. Il soggetto che progetta, addestra, commercializza o utilizza un sistema può non conoscere integralmente i meccanismi attraverso i quali esso giungerà a un determinato risultato. Da tale circostanza non discende automaticamente l'assenza di colpa: l'opacità del sistema può essere, entro certi limiti, una caratteristica conosciuta o conoscibile, e può dunque costituire proprio uno

[1] Sul problema della distribuzione dei contributi causali nei sistemi complessi, cfr. D. Thompson, The Moral Responsibility of Public Officials: The Problem of Many Hands, in *American Political Science Review, 1980, p. 905 ss.; M. Bovens, The Quest for Responsibility: Accountability and Citizenship in Complex Organisations, Cambridge University Press, 1998.

[2] Per la ricostruzione del nesso causale e il ruolo della spiegazione scientifica nel processo penale, cfr. F. Stella, Leggi scientifiche e spiegazione causale nel diritto penale, 2ª ed., Giuffrè, 2000; R. Blaiotta, Causalità giuridica, in Cassazione penale, 2007; M. Donini, Il garantismo della condicio sine qua non e il prezzo del suo abbandono. Contributo all'analisi dei rapporti fra causalità e imputazione, in Rivista italiana di diritto e procedura penale, 2011, n. 2, p. 494 ss. La bibliografia della tesi richiama inoltre F. Viganò, Il rapporto di causalità nella giurisprudenza penale a dieci anni dalla sentenza Franzese, cit.

degli elementi che impongono cautele nella selezione dei dati, nella validazione del modello, nella definizione del contesto d'impiego e nella predisposizione di controlli umani adeguati. Tuttavia, neppure l'esistenza di un rischio tecnologico consente di trasformare la posizione ricoperta dall'agente in una fonte automatica di responsabilità. Occorre individuare la regola cautelare concretamente esigibile, il comportamento alternativo doveroso e la possibilità di muovere un rimprovero personale per l'esito verificatosi[3].

La crisi del paradigma dell'evento si manifesta, pertanto, anche sul piano temporale della tutela. Quando la responsabilità penale viene ricostruita esclusivamente a partire dalla verificazione di una lesione, il sistema interviene, almeno per quanto concerne quella specifica fattispecie, dopo che il rischio si è tradotto in un danno o in un altro evento tipico. In settori nei quali le decisioni automatizzate possono incidere sulla sicurezza delle persone, sull'integrità dei sistemi o su interessi collettivi di particolare rilievo, attendere la produzione dell'evento può apparire insufficiente rispetto alle esigenze di prevenzione. Il problema assume una rilevanza ancora maggiore quando il rischio deriva da una configurazione sistemica: l'assenza di misure di sicurezza, di procedure di verifica o di adeguati meccanismi di supervisione può esporre una pluralità indeterminata di soggetti a conseguenze potenzialmente gravi, anche se non si è ancora verificata una lesione individuale[4].

La riflessione sulla società del rischio consente di collocare questa difficoltà entro una trasformazione più ampia. Come evidenziato da Ulrich Beck, lo sviluppo tecnologico e industriale non produce soltanto benefici e nuove opportunità, ma contribuisce anche alla generazione di rischi che possono risultare difficili da localizzare, prevedere e governare mediante gli strumenti istituzionali tradizionali. Nel diritto penale, questa trasformazione sollecita una riflessione sui limiti di un modello di tutela prevalentemente costruito attorno alla lesione di beni giuridici già verificatasi. Il punto, tuttavia, non è identificare la società del rischio con una necessaria espansione della punibilità. Al contrario, proprio la difficoltà di prevedere e ricostruire gli effetti delle nuove tecnologie rende indispensabile interrogarsi sulle condizioni alle quali l'intervento penale anticipato possa considerarsi legittimo[5].

La tutela anticipata rappresenta, in questa prospettiva, una possibile risposta alla crisi del diritto penale d'evento, ma non una soluzione automaticamente preferibile. Il ricorso a fattispecie di pericolo può consentire di sanzionare condotte che generano un rischio penalmente significativo prima che questo si traduca in una lesione; al tempo stesso, può accentuare il rischio di una punizione fondata su valutazioni generiche di pericolosità, su

[3] Sul giudizio di colpa, l'esigibilità della condotta alternativa e l'individuazione della regola cautelare, cfr. A. Canepa, L'imputazione soggettiva della colpa, Giappichelli, 2011; D. Castronuovo, La colpa penale, Cedam, 2009; G. Marinucci, Innovazioni tecnologiche e scoperte scientifiche: costi e tempi di adeguamento delle regole di diligenza, in Rivista italiana di diritto e procedura penale, 2005.

[4] U. Beck, La società del rischio. Verso una seconda modernità, Carocci, 2000 (ed. orig. 1986); J.M. Silva Sánchez, L'espansione del diritto penale. Aspetti della politica criminale nelle società postindustriali, Giuffrè, 2004; F. Herzog, Società del rischio, diritto penale del rischio, regolazione del rischio, in L. Stortoni – L. Foffani (a cura di), Critica e giustificazione del diritto penale nel cambio di secolo, Giuffrè, 2004, p. 357 ss.

[5] C. Burchard, L'intelligenza artificiale come fine del diritto penale?, in Archivio penale, 2019; F. Basile, Intelligenza artificiale e diritto penale, in Rivista italiana di diritto e procedura penale, 2022; F. Basile, Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione, in Sistema penale, 2023.

obblighi indeterminati o sulla mera inosservanza di prescrizioni amministrative e tecniche. Per evitare tale esito, l'anticipazione della tutela deve poggiare su una descrizione sufficientemente determinata della condotta vietata, sull'individuazione del bene giuridico protetto e sulla definizione di una soglia di rischio coerente con i principi di offensività, proporzionalità e personalità della responsabilità penale. La violazione di una regola tecnica, per quanto rilevante, non può essere convertita senza ulteriori verifiche in un illecito penale: occorre stabilire perché quella violazione assuma una dimensione offensiva tale da giustificare la sanzione criminale[6].

È proprio in questa distinzione che si colloca la proposta di ripensare la tutela penale dei sistemi di intelligenza artificiale senza abbandonare le categorie fondamentali della responsabilità personale. Da un lato, il diritto penale d'evento conserva una funzione essenziale quando sia possibile accertare che un comportamento umano, colposo o doloso, abbia causalmente contribuito alla produzione di un risultato tipico. Dall'altro, nei casi in cui la pericolosità dipenda soprattutto dalla mancata adozione di misure di sicurezza, dalla carenza di supervisione o dall'impiego di sistemi in condizioni incompatibili con le cautele richieste, può essere opportuno valutare fattispecie di tutela anticipata, costruite attorno a condotte e soglie di rischio rigorosamente determinate. L'obiettivo non è sostituire la prova dell'evento con una presunzione di pericolosità, ma evitare che la tutela di beni giuridici fondamentali dipenda esclusivamente dalla possibilità di ricostruire *ex post* processi tecnologici opachi.

La crisi del diritto penale d'evento non coincide, dunque, con il suo superamento. Essa individua piuttosto il punto nel quale le categorie tradizionali sono chiamate a confrontarsi con processi di produzione del rischio che rendono più complessa l'attribuzione causale e soggettiva, senza per questo rendere recessivi i principi che delimitano la responsabilità penale. La sfida consiste nel preservare la funzione garantista dell'imputazione dell'evento e, contemporaneamente, nel verificare se, per specifiche condotte connesse alla gestione dei sistemi di intelligenza artificiale, una tutela anticipata possa offrire una risposta più adeguata, purché rispettosa dei principi di legalità, offensività e colpevolezza. È in questo equilibrio, e non nella rinuncia alle categorie penalistiche fondamentali, che può essere ricercata una risposta coerente alle trasformazioni della società tecnologica[7].

5. Responsabilità degli enti e valorizzazione dell'art. 8 del d.lgs. n. 231/2001

La complessità dei sistemi di intelligenza artificiale rende particolarmente evidente l'insufficienza di un modello di imputazione fondato esclusivamente sulla ricostruzione della condotta del singolo individuo, soprattutto quando l'evento lesivo costituisca l'esito

[6] C. Brusco, Rischio e pericolo, rischio consentito e principio di precauzione, cit.; F. Bricola, Teoria generale del reato, cit.; C. Perini, Il concetto di rischio nel diritto penale moderno, cit.

[7] Per la discussione penalistica sull'IA e sui limiti dell'imputazione tradizionale, cfr. A. Cappellini, Machina delinquere non potest?, in Studi sulla questione criminale, 2018; A. Cappellini, Reati colposi e tecnologie dell'intelligenza artificiale, in Archivio penale, 2021; Y. Bathaee, The Artificial Intelligence Black Box and the Failure of Intent and Causation, in Harvard Journal of Law & Technology, 2018; J. Burrell, How the Machine "Thinks", in Big Data & Society, 2016.

di una pluralità di decisioni, omissioni e scelte organizzative, ciascuna delle quali contribuisca, secondo modalità differenti, alla produzione del rischio. Il problema non riguarda soltanto la difficoltà di individuare materialmente l'autore della condotta causalmente rilevante, ma investe più radicalmente la possibilità di ricostruire, secondo le categorie tradizionali del diritto penale, la relazione tra un determinato comportamento umano e un evento verificatosi all'interno di un sistema sociotecnico caratterizzato dalla distribuzione delle competenze, dalla frammentazione dei poteri decisionali e dall'interazione tra componenti umane e artificiali. L'intelligenza artificiale, soprattutto nelle sue applicazioni più complesse, non si limita infatti a introdurre un nuovo fattore tecnologico all'interno di processi decisionali già esistenti, ma può modificare la struttura stessa di tali processi, rendendo più articolata l'individuazione dei centri di responsabilità e più problematica la ricostruzione delle condizioni nelle quali il rischio si è concretizzato in un evento penalmente rilevante.

Una simile configurazione può essere ricondotta, almeno in parte, al cosiddetto *too many hands problem*, espressione con la quale si identifica la difficoltà di attribuire una responsabilità individuale sufficientemente determinata quando il risultato dannoso derivi dall'interazione di numerosi soggetti, ciascuno titolare di competenze, informazioni e poteri decisionali soltanto parziali. Il problema, originariamente sviluppato nell'ambito della riflessione sulla responsabilità morale e politica delle organizzazioni pubbliche, assume una rilevanza peculiare nel contesto dell'impresa tecnologica, nella quale la progettazione, l'addestramento, la validazione, la commercializzazione e l'impiego operativo di un sistema di intelligenza artificiale possono essere affidati a soggetti differenti, collocati anche al di fuori della medesima struttura organizzativa. Il processo che conduce alla produzione dell'evento lesivo può così articolarsi in una successione di decisioni tecniche e organizzative che, considerate isolatamente, non appaiono necessariamente idonee a spiegare l'accadimento, ma che, valutate nella loro interazione, possono rivelare una complessiva inadeguatezza delle misure di prevenzione e controllo[8].

Si pensi, a titolo esemplificativo, all'impiego di un sistema algoritmico destinato a supportare decisioni potenzialmente incidenti sulla sicurezza delle persone. Il produttore potrebbe aver progettato il sistema sulla base di dati non sufficientemente rappresentativi; il fornitore potrebbe averne documentato in maniera incompleta i limiti operativi; l'impresa utilizzatrice potrebbe averlo integrato nei propri processi senza predisporre adeguate verifiche; i responsabili della sicurezza potrebbero non aver definito procedure di intervento in caso di risultati anomali; gli operatori potrebbero, infine, aver fatto affidamento sulle indicazioni generate dal sistema senza disporre delle competenze o delle informazioni necessarie per valutarne criticamente l'affidabilità. In una simile ipotesi, l'evento lesivo non sarebbe necessariamente riconducibile a un'unica decisione erronea, ma potrebbe rappresentare la concretizzazione di una sequenza di vulnerabilità, ciascuna delle quali

[8] Sul problema della responsabilità distribuita nei processi tecnologici e sull'opacità algoritmica, cfr. D. Thompson, The Moral Responsibility of Public Officials: The Problem of Many Hands, cit.; Y. Bathaee, The Artificial Intelligence Black Box and the Failure of Intent and Causation, cit.; I. Salvadori, Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale, cit.

avrebbe potuto essere intercettata attraverso un diverso presidio tecnico, organizzativo o umano.

La difficoltà dell'accertamento penale risiede, pertanto, nella necessità di distinguere il contributo causalmente rilevante di ciascun soggetto dalla mera partecipazione al processo produttivo, evitando tanto l'indebita frammentazione della responsabilità quanto la sua indiscriminata concentrazione sul soggetto che occupa la posizione gerarchicamente più elevata o che si trova più vicino, sul piano materiale, al momento dell'evento. La complessità organizzativa non può tradursi in una presunzione di responsabilità dei vertici, così come l'intervento di un sistema di intelligenza artificiale non può determinare, per il solo fatto della sua autonomia funzionale, l'automatica esclusione della responsabilità delle persone fisiche coinvolte. Resta necessario verificare, rispetto a ciascun soggetto, la titolarità di effettivi poteri impeditivi, la disponibilità delle informazioni rilevanti, la riconoscibilità del rischio, l'esigibilità della condotta alternativa e la sussistenza del necessario collegamento causale e soggettivo con l'evento, secondo la struttura della fattispecie incriminatrice applicabile[9].

È proprio in questo passaggio che la questione tecnologica si trasforma in un problema di teoria generale del reato. L'eventuale impossibilità di ricondurre l'evento alla condotta colposa di un singolo individuo non implica necessariamente che il rischio si sia prodotto al di fuori di qualsiasi sfera di controllo umano. Può accadere, al contrario, che il rischio sia stato generato o amplificato da un insieme di decisioni organizzative che hanno determinato le condizioni di impiego del sistema, selezionato i suoi obiettivi operativi, stabilito il livello di supervisione, definito le procedure di verifica e distribuito le responsabilità di intervento. In questi casi, la ricerca di una responsabilità esclusivamente individuale rischia di non cogliere adeguatamente la dimensione organizzativa del fatto, soprattutto quando le singole condotte risultino difficilmente isolabili dal contesto nel quale sono state assunte.

Ciò non significa, tuttavia, che debba essere superato il principio della responsabilità personale o che sia legittimo sostituire l'accertamento della colpa individuale con una generica valutazione negativa dell'organizzazione. Il problema consiste, piuttosto, nell'individuare gli strumenti attraverso i quali il diritto possa considerare la dimensione collettiva della produzione del rischio senza rinunciare ai principi di legalità, colpevolezza e personalità della responsabilità. La questione non è, dunque, se la complessità tecnologica giustifichi un abbassamento delle soglie di imputazione, bensì se le categorie tradizionali, applicate senza un'adeguata considerazione dei processi organizzativi, siano sufficienti a descrivere tutte le modalità attraverso le quali un rischio tecnologico può essere generato, governato e lasciato evolvere fino alla produzione dell'evento.

In questa prospettiva assume particolare rilievo la responsabilità da reato degli enti disciplinata dal d.lgs. 8 giugno 2001, n. 231, il cui impianto consente di spostare l'attenzione dalla sola condotta della persona fisica alla struttura organizzativa nell'ambito della quale

[9] Sulla necessità di accertare i poteri impeditivi, gli obblighi cautelari e il nesso tra condotta ed evento, cfr. A. Canepa, L'imputazione soggettiva della colpa, cit.; D. Castronuovo, La colpa penale, cit.; A. Gargani, Posizioni di garanzia nelle organizzazioni complesse, cit.

il reato presupposto è stato commesso. La rilevanza del modello non risiede nella possibilità di attribuire all'ente, in termini automatici, gli errori del sistema tecnologico utilizzato, bensì nella possibilità di accertare se la commissione del reato sia stata resa possibile da carenze organizzative riconducibili alla mancata predisposizione o all'inefficace attuazione di adeguati presidi di prevenzione. L'organizzazione cessa, in tal modo, di rappresentare un semplice sfondo fattuale rispetto alla condotta individuale e diviene un elemento centrale dell'accertamento della responsabilità dell'ente, nei limiti e secondo i criteri stabiliti dal decreto[10].

La categoria della colpa di organizzazione permette, in particolare, di interrogarsi sull'adeguatezza delle scelte attraverso le quali l'impresa ha strutturato il proprio rapporto con la tecnologia. In un contesto nel quale l'intelligenza artificiale interviene in processi produttivi, gestionali o decisionali, l'attenzione non può esaurirsi nella verifica dell'esistenza formale di procedure interne o di documenti di conformità. Occorre valutare se l'ente abbia effettivamente identificato i rischi connessi all'impiego del sistema, individuato le funzioni responsabili della sua supervisione, predisposto controlli proporzionati alla gravità delle conseguenze prevedibili, assicurato la tracciabilità delle decisioni rilevanti e definito procedure idonee a consentire l'intervento umano quando il funzionamento del sistema presenti anomalie o si discosti dalle condizioni per le quali era stato validato. A tali profili si aggiungono la qualità dei dati utilizzati, la verifica periodica delle prestazioni, la gestione degli aggiornamenti, la formazione del personale e la capacità dell'organizzazione di reagire alle informazioni che segnalino un mutamento del livello di rischio[11].

La specificità della colpa di organizzazione emerge, pertanto, dalla possibilità che il difetto prevenzionistico non coincida con una singola omissione individuale, ma si manifesti nella complessiva inadeguatezza dell'assetto attraverso il quale l'attività rischiosa è stata pianificata, distribuita e controllata. L'assenza di una procedura di verifica, la sovrapposizione delle competenze, l'indeterminatezza delle responsabilità di supervisione, l'insufficienza delle risorse destinate alla sicurezza o la mancata circolazione delle informazioni possono concorrere a determinare un contesto nel quale nessun soggetto dispone, isolatamente, degli strumenti necessari per prevenire l'evento. Ciò che rileva, ai fini dell'eventuale responsabilità dell'ente, non è però la mera constatazione che l'organizzazione avrebbe potuto essere migliore, ma l'accertamento di una carenza organizzativa giuridicamente significativa, valutata alla luce dei criteri normativi applicabili e del suo collegamento con la commissione del reato presupposto.

L'impiego di sistemi di intelligenza artificiale rende tale valutazione particolarmente complessa, poiché impone di considerare non soltanto i rischi conosciuti al momento

[10] D.lgs. 8 giugno 2001, n. 231, artt. 5, 6 e 7. Sul rapporto tra responsabilità individuale e responsabilità dell'ente, cfr. C.E. Paliero, Responsabilità per danno da prodotto tra persona fisica ed ente collettivo, in Rivista trimestrale di diritto penale dell'economia, 1990; A. Alessandri, Parte generale, in C. Pedrazzi et al. (a cura di), Manuale di diritto penale dell'impresa, 2ª ed., Monduzzi, 2000.

[11] Sul criterio della colpa di organizzazione e sui presupposti della responsabilità dell'ente, cfr. d.lgs. n. 231/2001, artt. 5-7; C.E. Paliero, L'autunno del patriarca, in Rivista italiana di diritto e procedura penale, 1994, p. 1241 ss.; M.A. Bartolucci, L'art. 8 d.lgs. 231/2001 nel triangolo di Penrose, in Diritto penale contemporaneo, 2017.

dell'adozione della tecnologia, ma anche le modalità attraverso le quali l'organizzazione ha gestito i limiti informativi e operativi del sistema durante il suo concreto utilizzo. Un modello formalmente adeguato al momento della sua introduzione può rivelarsi insufficiente qualora non siano previsti controlli successivi, procedure di aggiornamento o meccanismi di segnalazione delle anomalie. Simmetricamente, il verificarsi di un evento lesivo non dimostra, di per sé, che l'ente abbia omesso le cautele dovute: l'accertamento deve tenere conto delle conoscenze disponibili, dei rischi ragionevolmente prevedibili, delle misure esigibili nel caso concreto e dell'effettiva idoneità dei presidi organizzativi a prevenire la commissione del reato.

La colpa di organizzazione non può, dunque, essere trasformata in una responsabilità oggettiva per il rischio tecnologico, né può essere desunta esclusivamente ex post dal verificarsi dell'evento. Il suo accertamento richiede una valutazione che ricostruisca, secondo una prospettiva normativa e causale, le scelte organizzative effettivamente compiute e quelle che, in presenza dei presupposti richiesti dall'ordinamento, avrebbero dovuto essere adottate. Il punto decisivo è stabilire se il difetto dell'assetto organizzativo abbia assunto un significato penalmente rilevante in relazione al reato presupposto, e non semplicemente se, alla luce dell'esito dannoso, sia possibile individuare una misura che avrebbe potuto migliorare il funzionamento del sistema.

Questa impostazione consente di valorizzare il d.lgs. n. 231/2001 senza attribuirgli una funzione che il testo normativo non gli riconosce. La responsabilità dell'ente, infatti, non costituisce un generale strumento di compensazione di ogni lacuna dell'imputazione individuale, né può essere invocata ogni volta che l'impiego dell'intelligenza artificiale renda difficile identificare il responsabile del danno. Essa presuppone la commissione di un reato compreso nel catalogo dei reati presupposto, la sussistenza dei criteri di imputazione oggettiva previsti dal decreto e la verifica dei requisiti soggettivi e organizzativi applicabili alla fattispecie concreta. Il vantaggio del modello risiede nella sua capacità di rendere giuridicamente rilevante il contributo dell'organizzazione alla realizzazione del fatto, non nell'eliminazione dei presupposti che delimitano l'area della responsabilità.

In questo quadro, l'art. 8 del d.lgs. n. 231/2001 assume un significato particolarmente importante. La disposizione sancisce l'autonomia della responsabilità dell'ente, stabilendo che essa sussiste anche quando l'autore del reato non sia stato identificato o non sia imputabile, nonché quando il reato si estingua per una causa diversa dall'amnistia. La norma impedisce, entro il proprio ambito applicativo, che l'impossibilità di individuare una determinata persona fisica si traduca automaticamente nell'impossibilità di procedere nei confronti dell'ente. Essa riconosce, in altri termini, che l'accertamento della responsabilità dell'organizzazione non deve necessariamente dipendere dalla previa identificazione nominativa dell'autore materiale del reato, purché risultino dimostrati gli ulteriori elementi richiesti dalla disciplina[12].

[12] D.lgs. n. 231/2001, art. 8. Cfr. M.A. Bartolucci, L'art. 8 d.lgs. 231/2001 nel triangolo di Penrose, cit., sul rapporto tra autonomia della responsabilità dell'ente e accertamento del reato presupposto.

La portata della disposizione acquista una particolare rilevanza nel contesto dei sistemi di intelligenza artificiale, nei quali la distribuzione delle attività e delle competenze può rendere difficile individuare il soggetto cui ricondurre materialmente una determinata omissione o decisione. L'art. 8 consente di evitare che l'incertezza relativa all'identità dell'autore individuale costituisca, di per sé, un ostacolo insuperabile all'accertamento della responsabilità dell'ente. In questo senso, l'autonomia prevista dal decreto assume una funzione di continuità dell'accertamento: la mancata identificazione della persona fisica non determina necessariamente l'interruzione dell'indagine sulla possibile rilevanza penale delle carenze organizzative che abbiano reso possibile la commissione del reato.

È tuttavia necessario distinguere con precisione l'ipotesi dell'autore non identificato da quella nella quale non sia possibile dimostrare che un reato presupposto sia stato effettivamente commesso. L'art. 8 non consente di prescindere dall'accertamento del reato, né introduce una responsabilità dell'ente fondata sul solo verificarsi di un evento dannoso. La norma riguarda l'autonomia della responsabilità dell'organizzazione rispetto a determinate vicende della persona fisica, ma non elimina la necessità di accertare il fatto costituente reato e gli altri elementi costitutivi dell'illecito dell'ente. Ne consegue che l'opacità del sistema algoritmico, pur potendo rendere più difficoltosa la ricostruzione dell'accaduto, non può essere utilizzata per colmare lacune probatorie concernenti l'esistenza del reato, la sua riconducibilità al catalogo normativo o i criteri di imputazione all'organizzazione[13].

La distinzione è essenziale anche per evitare che la responsabilità dell'ente venga trasformata in un meccanismo di imputazione residuale, destinato a operare ogniqualvolta non sia possibile condannare una persona fisica. L'autonomia prevista dall'art. 8 non istituisce una responsabilità di risulta, né consente di considerare l'ente responsabile per il semplice fatto che l'indagine non sia riuscita a individuare l'autore materiale. Essa permette, piuttosto, di mantenere distinti due piani di accertamento che, pur comunicando tra loro, non coincidono integralmente: da un lato, la responsabilità della persona fisica, da verificare secondo le categorie del reato e i principi che ne governano l'imputazione; dall'altro, la responsabilità dell'ente, da accertare alla luce dei criteri specifici previsti dal d.lgs. n. 231/2001.

Il rapporto tra i due piani diviene particolarmente significativo quando l'evento lesivo sia maturato all'interno di una struttura organizzativa che abbia affidato a un sistema di intelligenza artificiale una parte rilevante del processo decisionale. In tali circostanze, l'indagine potrebbe non riuscire a ricostruire una condotta individuale dotata di tutti i requisiti necessari per l'affermazione della responsabilità penale di una determinata persona fisica, pur facendo emergere elementi relativi a carenze nella gestione del rischio, nella distribuzione delle competenze o nella predisposizione dei controlli. La possibilità di procedere nei confronti dell'ente, quando ne ricorrano tutti i presupposti normativi,

[13] D.lgs. n. 231/2001, artt. 2, 5-8. L'art. 8 disciplina l'autonomia della responsabilità dell'ente rispetto alla mancata identificazione o all'inimputabilità dell'autore del reato, ma non elimina la necessità di accertare il reato presupposto e gli altri elementi dell'illecito dell'ente.

consente di non confondere la mancata identificazione dell'autore con l'inesistenza di una responsabilità organizzativa giuridicamente accertabile.

Non si tratta, in altri termini, di sostituire alla responsabilità individuale una responsabilità collettiva indifferenziata, ma di riconoscere che l'organizzazione può assumere una propria rilevanza nell'ambito del sistema di imputazione delineato dal legislatore. La difficoltà di individuare il singolo responsabile non deve tradursi né nell'automatica impunità dell'ente né nell'automatica affermazione della sua responsabilità: deve, piuttosto, indurre a verificare se la vicenda concreta integri gli elementi dell'illecito previsto dal decreto, con particolare attenzione al ruolo svolto dall'assetto organizzativo nella commissione del reato presupposto.

Rimane, peraltro, un limite strutturale che merita di essere evidenziato. L'art. 8 presuppone pur sempre l'esistenza di un reato presupposto e, pertanto, non risolve direttamente le ipotesi nelle quali l'intervento dell'intelligenza artificiale renda impossibile dimostrare la commissione di un reato riconducibile a una fattispecie incriminatrice vigente. Il suo potenziale applicativo non deve quindi essere sovrastimato: la disposizione può attenuare le conseguenze della mancata identificazione dell'autore, ma non può colmare eventuali vuoti di tipicità, superare l'assenza di prova degli elementi costitutivi del reato o consentire l'estensione analogica del catalogo dei reati presupposto.

Proprio questa delimitazione permette di cogliere il significato sistematico della responsabilità da reato degli enti nell'ambito del rischio tecnologico. Il d.lgs. n. 231/2001 offre una prospettiva attraverso la quale indagare la dimensione organizzativa della produzione del rischio, ma non può essere considerato, allo stato della disciplina, una soluzione generale a ogni possibile *responsibility gap* generato dall'intelligenza artificiale. Il problema che emerge è più ampio e riguarda la capacità dell'ordinamento di individuare forme di responsabilità coerenti con processi nei quali la produzione dell'evento non coincide necessariamente con una singola decisione umana, senza rinunciare alle garanzie che delimitano l'intervento penale[14].

La sfida consiste, pertanto, nel coniugare l'esigenza di evitare che la frammentazione organizzativa e tecnologica produca aree di irresponsabilità con la necessità di impedire che la complessità del sistema divenga un argomento per ampliare indiscriminatamente l'imputazione. In questa prospettiva, la responsabilità dell'ente può rappresentare uno strumento importante per valutare le carenze nella gestione del rischio algoritmico, a condizione che l'accertamento rimanga ancorato alla commissione di un reato presupposto e ai criteri di imputazione stabiliti dalla legge. La tecnologia non determina, dunque, il superamento dei principi fondamentali del diritto penale; rende piuttosto più urgente una loro applicazione capace di considerare la dimensione organizzativa dell'azione e le nuove modalità attraverso le quali il rischio viene prodotto, distribuito e governato.

[14] Cfr. C. Burchard, *L'intelligenza artificiale come fine del diritto penale?*, cit.; F. Basile, *Intelligenza artificiale e diritto penale*, cit.; A. Cappellini, *Reati colposi e tecnologie dell'intelligenza artificiale*, cit.

6. Il d.lgs. 9 settembre 2026, n. 160: conferme normative e limiti della nuova disciplina

Le questioni teoriche fin qui esaminate trovano un riscontro significativo nel d.lgs. 9 settembre 2026, n. 160, adottato in attuazione della delega conferita dalla legge 23 settembre 2025, n. 132, e destinato a incidere, tra gli altri profili, sulla disciplina penalistica della realizzazione e dell'impiego illecito dei sistemi di intelligenza artificiale. Il provvedimento si inserisce nel processo di adeguamento dell'ordinamento nazionale al regolamento (UE) 2024/1689, noto come *Artificial Intelligence Act*, ma presenta una portata che non si esaurisce nel coordinamento della disciplina interna con quella europea. Attraverso l'introduzione di nuove fattispecie incriminatrici e l'ampliamento del catalogo dei reati presupposto della responsabilità degli enti, il legislatore interviene direttamente sul rapporto tra rischio tecnologico, obblighi di prevenzione e imputazione della responsabilità penale, riconoscendo la necessità di predisporre strumenti di tutela specificamente calibrati sulle caratteristiche dei sistemi di intelligenza artificiale[15].

La rilevanza sistematica della riforma deve essere apprezzata alla luce di una difficoltà che attraversa l'intera riflessione penalistica sull'intelligenza artificiale: la possibilità che il rischio non sia riconducibile a una singola condotta isolata, ma rappresenti il risultato di una pluralità di decisioni, omissioni e scelte organizzative distribuite lungo l'intero ciclo di vita del sistema. La progettazione, la selezione e il trattamento dei dati, l'addestramento del modello, la validazione delle prestazioni, l'integrazione nei processi produttivi, la definizione delle procedure di supervisione e l'impiego operativo possono essere affidati a soggetti differenti, titolari di competenze e poteri non necessariamente coincidenti. In tale contesto, il verificarsi di un evento lesivo può dipendere dall'interazione tra fattori tecnici, comportamenti umani e carenze organizzative, rendendo particolarmente complessa l'individuazione del contributo individuale penalmente rilevante.

La questione non consiste, tuttavia, nell'affermare che l'impiego dell'intelligenza artificiale renda impossibile l'imputazione individuale. Una simile conclusione sarebbe eccessiva e finirebbe per attribuire alla complessità tecnologica un significato esonerante che non le appartiene. Il problema consiste piuttosto nel verificare se gli strumenti tradizionali dell'imputazione siano in grado di ricostruire adeguatamente le responsabilità quando il processo causale attraversi una struttura organizzativa articolata e quando le informazioni necessarie a comprendere il funzionamento del sistema siano distribuite tra una pluralità di soggetti. La difficoltà di individuare un unico centro decisionale non esclude, infatti, che siano identificabili specifiche condotte colpose o dolose, ma impone un accertamento particolarmente rigoroso dei doveri gravanti su ciascun soggetto, dei poteri effettivamente esercitabili, delle informazioni disponibili e del contributo della condotta alla realizzazione del fatto.

[15] D.lgs. 9 settembre 2026, n. 160, art. 12; legge 23 settembre 2025, n. 132; regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale (AI Act). Riferimenti normativi da controllare sul testo ufficiale prima della pubblicazione.

È in questa prospettiva che assume rilievo l'introduzione, ad opera dell'art. 12 del d.lgs. n. 160/2026, dell'art. 437-bis c.p., rubricato «Omessa adozione di misure di sicurezza nei sistemi di intelligenza artificiale e alterazione illecita dei sistemi». La disposizione introduce una pluralità di fattispecie concernenti l'omissione delle misure tecniche di sicurezza, la mancata adozione di misure di sorveglianza umana e determinate condotte di alterazione illecita dei sistemi. La prima fattispecie, in particolare, sanziona l'omessa adozione delle misure tecniche di sicurezza previste per la progettazione, l'addestramento, la produzione e l'immissione sul mercato di sistemi di intelligenza artificiale ad alto rischio, idonee a prevenire malfunzionamenti o alterazioni del funzionamento, nonché l'omessa adozione di misure di sorveglianza umana, quando da tali omissioni derivi pericolo per la vita o l'incolumità pubblica o individuale[16].

La scelta di collocare la nuova disposizione nell'ambito dei delitti contro l'incolumità pubblica rivela la volontà di attribuire rilievo penale a condotte che possono compromettere la sicurezza dei sistemi prima ancora che si produca una lesione effettiva. Il disvalore della condotta non viene dunque individuato esclusivamente nell'eventuale danno finale, ma anche nella creazione o nell'incremento di un rischio per beni giuridici fondamentali attraverso la violazione di specifici obblighi di sicurezza e controllo. La disciplina si inserisce, sotto questo profilo, nella più ampia evoluzione del diritto penale contemporaneo, nel quale la tutela anticipata dei beni giuridici assume una funzione crescente in presenza di attività tecnologiche capaci di generare conseguenze gravi e difficilmente reversibili.

Occorre, tuttavia, evitare di assimilare senza ulteriori precisazioni la nuova fattispecie al modello tradizionale del diritto penale d'evento. Quest'ultimo, nella sua configurazione classica, presuppone l'accertamento di un evento naturalistico e del rapporto causale che lo collega alla condotta, mentre il delitto di cui all'art. 437-bis c.p. presenta una struttura incentrata sull'omissione di determinate misure e sulla derivazione da essa di una situazione di pericolo. Il suo interesse per la riflessione sull'intelligenza artificiale risiede proprio nel fatto che la tutela penale può intervenire senza richiedere necessariamente la verificazione di una lesione, ma non per questo prescinde dalla dimostrazione degli elementi costitutivi della fattispecie. L'accertamento dovrà riguardare l'esistenza dell'obbligo di sicurezza o sorveglianza, la condotta omissiva, il pericolo richiesto dalla disposizione e il collegamento tra l'omissione e tale pericolo, oltre agli ulteriori requisiti soggettivi previsti dalla disciplina applicabile.

Quest'ultimo profilo merita particolare attenzione. Il testo definitivo della disposizione utilizza il riferimento al «pericolo», senza riprodurre l'espressione «pericolo concreto» presente nella formulazione dello schema originario. La scelta lessicale non può essere considerata irrilevante sul piano interpretativo, soprattutto in una materia nella quale la determinazione delle condizioni di punibilità dipende anche dall'individuazione delle caratteristiche del rischio che il legislatore intende prevenire. La formulazione della norma richiede, pertanto, un'attenta ricostruzione dei suoi elementi costitutivi, senza poter presumere che il riferimento al pericolo elimini ogni necessità di dimostrare la rilevanza

[16] D.lgs. 9 settembre 2026, n. 160, art. 12, in relazione all'introduzione dell'art. 437-bis c.p. Anche in questo caso è opportuno verificare il testo definitivo e la sua esatta rubrica sulla Gazzetta Ufficiale.

effettiva della situazione generata dalla condotta. Al contempo, non sarebbe corretto attribuire automaticamente alla fattispecie una struttura di pericolo concreto sulla sola base della funzione di tutela perseguita: la questione deve essere affrontata attraverso l'interpretazione del testo normativo e il confronto con i principi generali in materia di reati di pericolo[17].

Un'ulteriore difficoltà deriva dal rinvio alle misure tecniche di sicurezza previste per i sistemi di intelligenza artificiale ad alto rischio e alle misure di sorveglianza umana. Il contenuto degli obblighi la cui violazione può assumere rilevanza penale non può essere determinato esclusivamente attraverso valutazioni tecniche formulate dopo il verificarsi dell'evento. Occorre individuare, sulla base della normativa applicabile e delle caratteristiche del sistema, quali misure fossero effettivamente dovute, quali soggetti fossero tenuti ad adottarle e in quali condizioni la loro omissione possa essere ricondotta alla fattispecie incriminatrice. Il collegamento con la disciplina europea sull'intelligenza artificiale rende questa operazione particolarmente delicata, poiché richiede di coordinare obblighi regolatori di natura tecnica e organizzativa con le esigenze di determinatezza proprie della norma penale.

La questione si manifesta con particolare evidenza quando le diverse fasi del ciclo di vita del sistema siano affidate a soggetti distinti. Il produttore potrebbe essere responsabile della progettazione e della predisposizione delle misure di sicurezza; il fornitore potrebbe intervenire nell'integrazione del modello in un prodotto o servizio; l'impresa utilizzatrice potrebbe essere tenuta a predisporre procedure di supervisione adeguate al contesto operativo. La distribuzione delle funzioni non consente di trasferire indistintamente su tutti i soggetti coinvolti il medesimo obbligo di garanzia, né di presumere che ciascuno disponga delle conoscenze e dei poteri necessari per impedire ogni malfunzionamento. La responsabilità deve essere ricostruita in relazione alla posizione concretamente rivestita, agli obblighi applicabili e alla possibilità effettiva di adempiervi.

Il rischio è che la complessità del sistema venga utilizzata in due direzioni opposte, entrambe problematiche. Da un lato, la frammentazione delle competenze potrebbe favorire una dispersione della responsabilità, nella quale ciascun soggetto attribuisca ad altri il compito di prevenire il rischio. Dall'altro, la gravità dell'evento potrebbe indurre a concentrare la responsabilità sul soggetto più facilmente individuabile, anche quando non sia dimostrato che questi disponesse dei poteri o delle informazioni necessari a impedire il fatto. La nuova incriminazione non elimina tale tensione, ma rende più urgente una ricostruzione rigorosa dei doveri di sicurezza e dei rapporti tra le diverse posizioni organizzative.

È proprio in questo passaggio che il collegamento con il cosiddetto *too many hands problem* acquista un significato ulteriore. Il problema delle molte mani non coincide semplicemente

[17] Sul rapporto tra reati di pericolo, offensività e anticipazione della tutela, cfr. F. Bricola, Teoria generale del reato, in Novissimo Digesto Italiano, 1973; C. Brusco, Rischio e pericolo, rischio consentito e principio di precauzione, in Cassazione penale, 2006; C. Perini, Il concetto di rischio nel diritto penale moderno, Giuffrè, 2010.

con l'esistenza di numerosi partecipanti al processo tecnologico: consiste nella difficoltà di ricostruire una responsabilità individuale sufficientemente determinata quando il risultato dannoso dipenda dall'interazione di contributi parziali, decisioni distribuite e omissioni reciprocamente condizionate. La previsione di specifici obblighi di sicurezza può contribuire a rendere più riconoscibili le condotte doverose e a delimitare le aree di responsabilità, ma non risolve automaticamente la difficoltà di stabilire chi fosse tenuto ad adottare una determinata misura e se la sua omissione sia penalmente rilevante nel caso concreto.

L'intervento legislativo assume un significato ancora più rilevante se considerato in connessione con l'art. 15 del d.lgs. n. 160/2026, che introduce nel d.lgs. n. 231/2001 l'art. 25-vicies, rubricato «Reati commessi con l'uso di sistemi di intelligenza artificiale». La disposizione inserisce nel catalogo dei reati presupposto il nuovo art. 437-bis c.p. e il delitto di cui all'art. 612-quater c.p., relativo all'illecita diffusione di contenuti generati o manipolati artificialmente. La riforma riconosce così espressamente che determinate condotte illecite connesse all'intelligenza artificiale possono fondare, in presenza degli ulteriori presupposti normativi, una responsabilità dell'ente distinta da quella della persona fisica che abbia commesso il reato[18].

Il significato di questa scelta non deve essere ridotto a un semplice ampliamento quantitativo del catalogo dei reati presupposto. L'inserimento delle nuove fattispecie consente di considerare la dimensione organizzativa di condotte che possono essere favorite da assetti aziendali inadeguati, da procedure di controllo insufficienti o da una distribuzione delle competenze incapace di assicurare un'effettiva gestione del rischio. Nel caso dell'art. 437-bis c.p., l'indagine potrà riguardare l'adeguatezza delle procedure di sicurezza, l'effettività della sorveglianza umana, la definizione delle responsabilità interne e la predisposizione di controlli idonei a prevenire le omissioni penalmente rilevanti. Nel caso dell'art. 612-quater c.p., potranno invece assumere rilievo le procedure attraverso le quali l'ente governa la generazione, la verifica e la diffusione di contenuti artificialmente prodotti o manipolati, tenendo conto della natura dell'attività svolta e delle circostanze concrete della commissione del reato.

La responsabilità dell'ente, tuttavia, non può essere dedotta dalla sola circostanza che il reato sia stato commesso mediante un sistema di intelligenza artificiale. L'art. 25-vicies amplia le fattispecie che possono costituire il presupposto dell'illecito, ma non trasforma l'impiego della tecnologia in un criterio autonomo e sufficiente di imputazione. Rimangono necessari i requisiti stabiliti dal d.lgs. n. 231/2001, a partire dalla commissione di un reato presupposto e dalla sua riconducibilità all'ente secondo i criteri previsti dagli artt. 5, 6 e 7, secondo la disciplina applicabile. Occorre inoltre distinguere la posizione dell'ente da quella della persona fisica: l'eventuale responsabilità individuale non determina automaticamente quella dell'organizzazione, così come la difficoltà di identificare l'autore materiale non

[18] D.lgs. 8 giugno 2001, n. 231, artt. 5-8 e art. 25-vicies, come richiamato nel testo; d.lgs. 9 settembre 2026, n. 160, art. 15. Verificare sul testo ufficiale l'esatta numerazione e il contenuto delle disposizioni introdotte.

esclude, di per sé, l'accertamento della responsabilità dell'ente quando ne ricorrano tutti i presupposti.

La categoria della colpa di organizzazione assume, in questo contesto, una funzione centrale. Essa permette di interrogarsi non soltanto sulla condotta del singolo che abbia materialmente violato un obbligo, ma anche sull'adeguatezza dell'assetto attraverso il quale l'impresa ha organizzato l'attività rischiosa. La distribuzione delle competenze, la definizione dei poteri decisionali, la circolazione delle informazioni, la formazione del personale, la verifica periodica del funzionamento dei sistemi e la predisposizione di procedure di intervento possono concorrere a determinare il livello di controllo effettivamente esercitato dall'organizzazione sulla tecnologia impiegata.

L'analisi deve però evitare una sovrapposizione tra colpa individuale e colpa di organizzazione. La prima riguarda la violazione del dovere cautelare da parte della persona fisica, secondo i requisiti propri della fattispecie contestata; la seconda concerne l'eventuale deficit dell'organizzazione rilevante ai fini dell'illecito dell'ente. Le due dimensioni possono interagire, ma non coincidono. Una condotta individuale negligente può verificarsi anche all'interno di un'organizzazione adeguatamente strutturata, così come un assetto organizzativo inadeguato può contribuire alla commissione del reato attraverso una pluralità di decisioni e omissioni che non si lasciano ricondurre agevolmente a un unico soggetto. La responsabilità dell'ente richiede, pertanto, una valutazione specifica della struttura organizzativa e del suo rapporto con il reato presupposto, senza che la mera verificazione dell'illecito individuale sia sufficiente a dimostrare il difetto organizzativo.

In tale prospettiva, l'adozione di un modello di organizzazione, gestione e controllo non può essere valutata sulla base della sola esistenza formale di documenti o procedure. Occorre interrogarsi sulla sua concreta idoneità preventiva, sulla capacità di individuare i rischi derivanti dall'impiego dei sistemi di intelligenza artificiale e sull'effettività dei meccanismi di controllo. Un modello che preveda genericamente l'obbligo di supervisione umana, senza individuare chi debba esercitarla, con quali strumenti, secondo quali procedure e in presenza di quali anomalie, potrebbe risultare insufficiente rispetto ai rischi specifici dell'attività. Analogamente, la previsione di controlli tecnici non accompagnati da un sistema di segnalazione, aggiornamento e intervento potrebbe non garantire un'effettiva capacità di prevenzione.

La valutazione deve essere condotta senza trasformare la colpa di organizzazione in una forma di responsabilità oggettiva. Il verificarsi di un evento lesivo o la commissione di un reato presupposto non dimostrano, da soli, che l'ente abbia adottato un'organizzazione inadeguata. È necessario ricostruire le misure effettivamente adottate, le carenze eventualmente riscontrate, i rischi conoscibili al momento delle decisioni e il significato di tali carenze rispetto alla commissione del reato. L'accertamento non può essere costruito esclusivamente ex post, attraverso la selezione di tutte le misure che, una volta verificatosi l'evento, appaiano astrattamente utili a evitarlo.

La riforma assume così un significato che va oltre la mera introduzione di nuovi reati. Essa rende più evidente l'esigenza di una governance del rischio algoritmico capace di tradurre

le prescrizioni normative in procedure effettive, verificabili e coerenti con le caratteristiche della tecnologia. L'organizzazione non può limitarsi a dichiarare la conformità del sistema agli obblighi applicabili, ma deve essere in grado di dimostrare, nei limiti delle responsabilità e degli obblighi che le competono, di aver predisposto misure adeguate alla prevenzione dei rischi rilevanti. Al contempo, la conformità regolatoria non può essere considerata una garanzia assoluta di assenza di responsabilità penale: essa rappresenta un elemento da valutare nel quadro complessivo degli obblighi applicabili, senza sostituire l'accertamento degli elementi costitutivi del reato o dell'illecito dell'ente.

Resta, infine, il problema del rapporto tra la nuova disciplina e l'autonomia della responsabilità dell'ente sancita dall'art. 8 del d.lgs. n. 231/2001. La disposizione consente di mantenere ferma la responsabilità dell'organizzazione anche quando l'autore del reato non sia stato identificato o non sia imputabile, nei limiti previsti dalla norma. In un contesto caratterizzato dalla frammentazione dei processi decisionali e dalla difficoltà di ricostruire il contributo individuale, tale autonomia può assumere una funzione importante, impedendo che la mancata identificazione della persona fisica determini automaticamente l'impossibilità di procedere nei confronti dell'ente.

Anche sotto questo profilo, tuttavia, occorre evitare conclusioni eccessivamente estensive. L'art. 8 non consente di prescindere dalla dimostrazione del reato presupposto, né permette di desumere la responsabilità dell'ente dal solo verificarsi di un evento dannoso. La mancata identificazione dell'autore non equivale all'impossibilità di provare il reato, ma neppure autorizza a considerarlo esistente quando gli elementi necessari non siano stati accertati. L'autonomia dell'illecito dell'ente opera entro la struttura normativa del decreto e non costituisce uno strumento per colmare eventuali lacune probatorie relative alla tipicità del fatto, alla sua commissione o ai criteri di imputazione dell'organizzazione.

Il coordinamento tra l'art. 437-bis c.p., l'art. 25-vicies e l'art. 8 del d.lgs. n. 231/2001 consente pertanto di individuare un modello di tutela articolato su piani distinti ma complementari. La nuova fattispecie incriminatrice individua specifiche condotte che compromettono la sicurezza dei sistemi di intelligenza artificiale; l'estensione del catalogo dei reati presupposto permette di valutare l'eventuale responsabilità dell'ente per la commissione di tali reati; l'autonomia prevista dall'art. 8 impedisce, infine, che determinate vicende processuali della persona fisica precludano automaticamente l'accertamento relativo all'organizzazione. Nessuna di queste disposizioni, isolatamente considerata, risolve il problema generale dell'imputazione del rischio algoritmico, ma la loro interazione contribuisce a delineare gli strumenti attraverso i quali l'ordinamento può affrontarlo.

La portata innovativa della riforma deve quindi essere misurata non soltanto in relazione all'ampliamento delle fattispecie penalmente rilevanti, ma anche alla capacità del sistema di individuare responsabilità effettive senza sacrificare le garanzie fondamentali. Il diritto penale è chiamato a evitare che la complessità tecnologica e organizzativa produca aree di irresponsabilità, ma deve al contempo impedire che tale complessità diventi una giustificazione per estendere indiscriminatamente l'imputazione. La responsabilità non può essere trasferita automaticamente sul produttore, sull'utilizzatore, sui vertici aziendali o sull'ente nel suo complesso soltanto perché il sistema ha generato un risultato dannoso o

pericoloso. Deve essere ricostruita alla luce dei doveri concretamente gravanti sui soggetti, delle possibilità di controllo effettivamente disponibili e dei requisiti normativi richiesti per ciascuna forma di responsabilità.

In definitiva, la nuova disciplina conferma che il problema dell'intelligenza artificiale non consiste semplicemente nell'individuare un soggetto cui attribuire gli effetti negativi della tecnologia, ma nel definire criteri di imputazione capaci di distinguere la mera partecipazione al processo tecnologico dalla responsabilità giuridicamente rilevante. L'ampliamento del catalogo dei reati presupposto e la valorizzazione degli obblighi di sicurezza possono contribuire a rendere più effettiva la tutela penale, ma non sostituiscono la necessità di dimostrare gli elementi costitutivi dell'illecito. La sfida consiste, pertanto, nel coniugare la prevenzione dei rischi algoritmici con la determinatezza delle fattispecie, la responsabilizzazione delle organizzazioni con il divieto di responsabilità oggettiva e la tutela dei beni giuridici con il rispetto dei principi di legalità e colpevolezza. È in questo equilibrio che si misura la capacità del diritto penale di confrontarsi con le trasformazioni tecnologiche senza rinunciare alla propria struttura garantista.

7. Conclusioni: nuove soluzioni per nuovi problemi o vecchie soluzioni per vecchi problemi?

La sfida del diritto penale dell'intelligenza artificiale non consiste nel punire la macchina, né nel trasferire automaticamente sull'essere umano ogni conseguenza dannosa del suo funzionamento. Una simile alternativa, apparentemente esaustiva, rischia infatti di semplificare eccessivamente un problema che investe alcuni dei fondamenti della responsabilità penale: l'individuazione del soggetto cui ricondurre il fatto, la ricostruzione del nesso causale, la determinazione dei doveri di condotta e l'accertamento della rimproverabilità personale. La questione non riguarda soltanto la capacità del diritto penale di reagire a forme inedite di pericolo, ma investe la sua stessa capacità di conservare una struttura garantista di fronte a processi decisionali nei quali l'intervento umano può risultare frammentato, mediato da strumenti tecnologici e distribuito lungo una pluralità di fasi organizzative.

L'intelligenza artificiale non costituisce, sotto questo profilo, un fenomeno capace di determinare automaticamente il superamento delle categorie tradizionali del reato. Essa rappresenta, piuttosto, un banco di prova della loro effettiva capacità esplicativa. La circostanza che un sistema possa elaborare informazioni, produrre risultati non predeterminati in ogni dettaglio o modificare le proprie prestazioni in relazione ai dati ricevuti non comporta, di per sé, che il diritto penale debba rinunciare alle proprie categorie fondamentali. Impone, invece, di verificare in quali condizioni tali categorie siano applicabili, quali difficoltà interpretative emergano dalla trasformazione tecnologica e se gli strumenti normativi disponibili siano sufficienti a governare le nuove modalità di produzione e distribuzione del rischio. La novità tecnologica, dunque, non costituisce una ragione sufficiente per abbandonare il sistema tradizionale dell'imputazione, così come il

richiamo alla continuità dei principi non può diventare un argomento per ignorare le trasformazioni della realtà che tali principi sono chiamati a disciplinare.

Il problema fondamentale risiede nella progressiva dissociazione tra il luogo nel quale viene assunta una decisione, il soggetto che dispone delle informazioni necessarie per comprenderla e quello che possiede i poteri per impedirne le conseguenze. Nei sistemi di intelligenza artificiale impiegati in contesti complessi, la produzione del rischio può dipendere da scelte compiute in momenti differenti e da soggetti che operano in organizzazioni distinte. Il produttore può definire l'architettura del sistema, il fornitore può integrarlo in un determinato ambiente operativo, l'utilizzatore può determinarne le condizioni concrete di impiego e l'organizzazione può stabilire il livello di supervisione e le procedure di intervento. L'evento lesivo può così emergere dall'interazione tra decisioni che, considerate isolatamente, non appaiono sufficienti a spiegare l'accadimento, ma che, valutate nel loro complesso, possono rivelare un'insufficiente gestione del rischio.

Una simile configurazione mette in discussione l'idea, talvolta implicitamente presupposta nei modelli tradizionali di ricostruzione dell'illecito, che sia sempre possibile individuare una condotta individuale chiaramente separabile dal contesto nel quale essa è stata realizzata. La difficoltà non deriva necessariamente dall'assenza di un contributo umano, ma può dipendere dalla distribuzione delle conoscenze, dalla frammentazione delle competenze e dalla distanza tra le decisioni organizzative e il momento in cui il rischio si concretizza. Il problema del *too many hands* assume, in questo senso, una rilevanza propriamente penalistica: non basta constatare che numerosi soggetti abbiano partecipato al processo tecnologico, ma occorre stabilire quale contributo sia giuridicamente rilevante, quali obblighi gravassero su ciascuno e se la condotta individuale soddisfi tutti gli elementi richiesti dalla fattispecie incriminatrice.

La risposta non può consistere nell'attribuzione indifferenziata della responsabilità a tutti coloro che abbiano preso parte al processo. Una simile soluzione, pur potendo apparire funzionale all'esigenza di assicurare una tutela effettiva, finirebbe per sostituire all'accertamento della responsabilità personale una logica di imputazione fondata sulla mera partecipazione all'attività rischiosa. Il coinvolgimento nella progettazione, nella commercializzazione o nell'utilizzazione di un sistema non equivale, infatti, alla titolarità di un obbligo giuridico di impedire qualsiasi conseguenza dannosa. È necessario individuare la posizione concretamente rivestita, i poteri esercitabili, le informazioni disponibili, la prevedibilità del rischio e le misure esigibili nel caso specifico. Soltanto attraverso tale ricostruzione è possibile distinguere la responsabilità penalmente rilevante dalla semplice appartenenza a una catena causale, tecnica o organizzativa.

In questo quadro, il principio di causalità conserva una funzione imprescindibile, ma incontra difficoltà applicative che non possono essere sottovalutate. Quando l'evento sia il risultato di una pluralità di fattori, la ricostruzione del nesso causale deve evitare tanto l'attribuzione automatica del risultato a una condotta temporalmente o materialmente vicina all'evento quanto l'esclusione della responsabilità per il solo fatto che il processo sia

stato mediato da un sistema algoritmico. L'impiego dell'intelligenza artificiale non elimina la necessità di accertare il rapporto tra condotta ed evento nei reati che lo richiedono, né autorizza a sostituire la prova causale con una valutazione generale sull'inadeguatezza della gestione tecnologica. La complessità del processo può rendere l'accertamento più difficile, ma non modifica, di per sé, lo standard probatorio necessario per affermare la responsabilità penale.

La medesima cautela deve caratterizzare l'accertamento della colpevolezza. La disponibilità di un sistema tecnologico avanzato non implica che ogni conseguenza del suo funzionamento fosse prevedibile, così come l'affidamento su un sistema automatizzato non esclude necessariamente la possibilità di un rimprovero. Occorre verificare se il rischio fosse riconoscibile alla luce delle conoscenze disponibili, se il soggetto fosse tenuto a svolgere determinate verifiche, se disponesse degli strumenti necessari per farlo e se fosse concretamente esigibile una condotta alternativa. L'eventuale imprevedibilità di una specifica modalità di funzionamento, l'insufficienza delle informazioni accessibili o l'assenza di effettivi poteri di intervento possono assumere rilievo nell'accertamento della responsabilità, ma non possono essere valutati in astratto: la loro incidenza dipende dalla posizione del soggetto, dalle caratteristiche dell'attività e dal contenuto degli obblighi giuridici applicabili.

Proprio in questo passaggio emerge la differenza tra la necessità di aggiornare le categorie penalistiche e la tentazione di superarle. L'adeguamento del diritto penale alla realtà tecnologica può richiedere una più precisa definizione degli obblighi di sicurezza, una migliore individuazione delle posizioni di garanzia e l'introduzione di fattispecie specificamente destinate a prevenire rischi che non trovino una risposta adeguata nella disciplina vigente. Non richiede, invece, che la causalità venga sostituita dalla mera connessione tecnologica, che la colpevolezza sia desunta dal ruolo ricoperto nell'organizzazione o che la responsabilità personale venga trasformata in una responsabilità per il semplice verificarsi dell'evento. Il diritto penale può evolvere soltanto se conserva la distinzione tra il rischio che l'ordinamento intende prevenire e il fatto che può essere legittimamente imputato a un determinato soggetto.

La distinzione assume particolare rilievo nel rapporto tra diritto penale d'evento e anticipazione della tutela. L'impiego dell'intelligenza artificiale può determinare situazioni nelle quali l'intervento preventivo risulti necessario prima che si produca una lesione, soprattutto quando siano coinvolti beni giuridici quali la vita, l'incolumità personale e la sicurezza pubblica. In tali contesti, il legislatore può ritenere opportuno incriminare la violazione di specifici obblighi di sicurezza o condotte capaci di generare un pericolo penalmente rilevante. Tuttavia, la scelta di anticipare la tutela non elimina il problema della determinatezza della fattispecie, né consente di attribuire rilevanza penale a qualsiasi inosservanza di una regola tecnica. Occorre individuare con sufficiente precisione gli obblighi la cui violazione assume rilievo, i soggetti tenuti ad adempiervi e le condizioni nelle quali la condotta integra il fatto previsto dalla norma.

La prevenzione del rischio algoritmico deve pertanto essere distinta dalla punizione indiscriminata dell'incertezza tecnologica. Un sistema di intelligenza artificiale può

presentare limiti, margini di errore o comportamenti non pienamente prevedibili; tali caratteristiche non sono, da sole, sufficienti a fondare una responsabilità penale. Il diritto deve stabilire quali rischi possano essere considerati giuridicamente rilevanti, quali cautele siano esigibili e quando l'inosservanza di tali cautele superi la soglia necessaria per integrare un illecito. Diversamente, la mera presenza di una tecnologia complessa potrebbe diventare un fattore di aggravamento implicito della posizione di chi la utilizza, imponendo obblighi indefiniti e potenzialmente illimitati di controllo.

È proprio per evitare questa deriva che l'innovazione normativa deve essere valutata alla luce della sua effettiva necessità. L'introduzione di una nuova fattispecie è giustificata quando consente di colmare una lacuna reale di tutela, individua un disvalore sufficientemente determinato e risponde a un'esigenza che non può essere adeguatamente soddisfatta attraverso gli strumenti già disponibili. Non è sufficiente che una condotta sia tecnologicamente nuova perché debba essere necessariamente oggetto di una nuova incriminazione. Occorre verificare se essa sia già riconducibile a una fattispecie vigente, se sussistano effettive difficoltà applicative e se l'intervento legislativo sia proporzionato al rischio che intende prevenire. La novità della tecnologia, in altri termini, non coincide automaticamente con una lacuna dell'ordinamento.

Allo stesso tempo, l'esigenza di conservare le categorie tradizionali non deve tradursi in una loro applicazione meccanica, indifferente alle trasformazioni del contesto tecnologico. L'interpretazione delle regole cautelari, la ricostruzione delle posizioni di garanzia e la valutazione dell'esigibilità della condotta alternativa devono tenere conto delle modalità concrete con cui i sistemi vengono progettati, distribuiti e impiegati. La continuità dei principi non implica immobilità delle soluzioni interpretative. Implica, piuttosto, che ogni adattamento sia giustificato attraverso argomenti giuridici verificabili e rimanga entro i limiti consentiti dal testo normativo, senza ricorrere ad analogie *in malam partem* o a interpretazioni che rendano imprevedibile l'estensione della responsabilità penale.

Il medesimo equilibrio deve guidare la riflessione sulla responsabilità da reato degli enti. La complessità organizzativa dei processi di intelligenza artificiale rende particolarmente importante la possibilità di valutare le carenze dei sistemi di prevenzione e controllo, soprattutto quando il rischio sia stato prodotto o amplificato da una distribuzione inadeguata delle competenze, da procedure insufficienti o dall'assenza di una supervisione effettiva. La colpa di organizzazione offre, in questo senso, una prospettiva che consente di considerare la dimensione collettiva dell'attività senza dover necessariamente ricondurre ogni criticità a una singola decisione individuale. La sua funzione non è quella di colmare automaticamente l'eventuale impossibilità di individuare il responsabile, ma di accertare se l'ente abbia adottato un assetto organizzativo adeguato ai rischi rilevanti e se l'eventuale deficit abbia assunto significato rispetto alla commissione del reato presupposto.

L'ampliamento del catalogo dei reati presupposto può rafforzare la capacità dell'ordinamento di intervenire nei confronti di condotte illecite connesse all'impiego dell'intelligenza artificiale, ma non determina un'automatica responsabilità dell'organizzazione. La presenza di un reato presupposto, l'esistenza di un modello organizzativo e la concreta adeguatezza delle misure adottate devono essere valutate

secondo la disciplina applicabile, senza trasformare la responsabilità dell'ente in una forma di garanzia generale per tutti i rischi dell'attività tecnologica. Anche in questo ambito, l'effettività della tutela dipende dalla capacità di distinguere le carenze organizzative realmente rilevanti dalle imperfezioni che, pur potendo essere individuate a posteriori, non dimostrano necessariamente una violazione penalmente significativa.

La questione assume una dimensione ulteriore quando l'autore del reato non sia identificato. L'autonomia della responsabilità dell'ente prevista dall'art. 8 del d.lgs. n. 231/2001 impedisce che la mancata identificazione della persona fisica determini automaticamente l'esclusione della responsabilità dell'organizzazione, ma non consente di prescindere dalla dimostrazione del reato presupposto e degli altri elementi dell'illecito. Il sistema non autorizza, pertanto, a sostituire alla responsabilità individuale una responsabilità collettiva indifferenziata, né a considerare l'ente responsabile per il solo fatto che l'indagine non sia riuscita a individuare l'autore materiale. L'autonomia opera entro i limiti normativi stabiliti e conferma che la complessità dell'organizzazione può richiedere una pluralità di piani di accertamento, ciascuno dotato di presupposti propri.

Ne deriva che il diritto penale dell'intelligenza artificiale non può essere costruito intorno alla ricerca di un responsabile a ogni costo, ma deve fondarsi sulla capacità di distinguere le diverse forme di rischio, i diversi livelli di intervento umano e i differenti criteri di imputazione. La persona fisica risponde quando il fatto le sia personalmente attribuibile secondo gli elementi della fattispecie applicabile; l'ente risponde quando ricorrano i presupposti specifici della responsabilità da reato; il sistema tecnologico, invece, costituisce l'oggetto e il contesto dell'attività rischiosa, non un espediente per eludere le condizioni che legittimano l'intervento punitivo. Questa impostazione non nega che la tecnologia possa modificare profondamente il modo in cui il rischio viene prodotto, ma impedisce che tale trasformazione venga utilizzata per dilatare indiscriminatamente l'area della punibilità.

La vera sfida consiste, dunque, nel mantenere aperto un duplice orizzonte. Da un lato, l'ordinamento deve essere capace di riconoscere i rischi nuovi e di intervenire quando le categorie esistenti non offrano una tutela adeguata; dall'altro, deve evitare che la necessità di prevenire tali rischi determini un progressivo arretramento delle garanzie. La prevenzione non può sostituire la prova, l'efficienza non può prevalere sulla determinatezza e la complessità tecnologica non può giustificare una responsabilità fondata sulla mera posizione occupata dal soggetto nel processo produttivo. Il compito del legislatore e dell'interprete consiste nel rendere le regole sufficientemente adeguate alle nuove forme di rischio senza sacrificare la prevedibilità delle conseguenze penali e la possibilità di riconoscere, nel caso concreto, il fondamento del rimprovero.

In definitiva, l'intelligenza artificiale non impone di scegliere tra un diritto penale incapace di comprendere la trasformazione tecnologica e un diritto penale disposto a rinunciare ai propri principi pur di assicurare una risposta a ogni evento dannoso. Tale alternativa è soltanto apparente. L'ordinamento può aggiornare le proprie fattispecie, precisare gli obblighi di sicurezza, valorizzare la dimensione organizzativa del rischio e sviluppare criteri interpretativi adeguati, mantenendo al contempo la causalità, la colpevolezza e la responsabilità personale quali condizioni fondamentali della legittimità della pena. La

modernizzazione del diritto penale non coincide con l'espansione indefinita del potere punitivo, ma con la capacità di individuare nuove soluzioni senza abbandonare le ragioni che ne delimitano l'esercizio.

La sfida del diritto penale dell'intelligenza artificiale non consiste, pertanto, nel costruire una responsabilità senza autore, né nel fare dell'essere umano il garante universale di ogni risultato prodotto da una macchina. Consiste nel definire un sistema di imputazione capace di riconoscere la complessità dei processi tecnologici, distinguere le responsabilità individuali da quelle organizzative e intervenire soltanto quando siano dimostrati i presupposti normativi del rimprovero. La tecnologia può modificare il modo in cui il rischio nasce, si distribuisce e si concretizza; non può, tuttavia, modificare la ragione fondamentale per cui il diritto penale punisce: non la mera produzione di un risultato indesiderato, ma la commissione di un fatto che l'ordinamento abbia previamente qualificato come reato e che possa essere personalmente rimproverato al suo autore secondo criteri legittimi e verificabili. È in questa continuità dei principi, accompagnata dalla capacità di adattarne rigorosamente l'applicazione, che deve essere ricercata la risposta più adeguata alle trasformazioni introdotte dall'intelligenza artificiale[19].

[19] Per la prospettiva generale della tesi, cfr. A. Cappellini, Machina delinquere non potest?, cit.; F. Basile, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione, cit.; sul rapporto tra causalità e imputazione soggettiva, F. Stella, Leggi scientifiche e spiegazione causale nel diritto penale, cit.