

One step forward to AI criminal liability. I nuovi rischi sanzionatori in materia di intelligenza artificiale nelle organizzazioni complesse: dall'AI Act ai decreti legislativi attuativi della "legge IA" (d.lgs. 160/2026 e d.lgs. 179/2026).

di **Matteo Riccardi** e **Marta Berardi***

Sommario. **1.** Premessa. Dall'AI Act alla normativa nazionale di adeguamento. – **2.** L'AI Act: impostazione metodologica e ambito di applicazione. – **2.1.** Ambito di applicazione oggettivo: la logica *risk-based* dei sistemi di intelligenza artificiale. – **2.2.** Ambito di applicazione soggettivo: la «catena del valore» dell'intelligenza artificiale. – **3.** L'AI Act: *framework* sanzionatorio. – **3.1.** Le misure sanzionatorie amministrative. – **4.** Intelligenza artificiale e diritto penale. Il delitto di «Omessa adozione di misure di sicurezza nei sistemi di intelligenza artificiale e alterazione illecita dei sistemi» (articolo 437-bis c.p.). – **4.1.** L'omessa adozione di misure di sicurezza e di sorveglianza umana per i sistemi di intelligenza artificiale ad alto rischio (articolo 437-bis, commi 1 e 3 c.p.). – **4.2.** L'alterazione di sistemi di intelligenza artificiale ad alto rischio (articolo 437-bis, comma 2, c.p.). – **4.3.** Responsabilità delle persone fisiche e *ne bis in idem* europeo. – **4.4.** Annotazioni conclusive sul delitto dell'articolo 437-bis c.p. – **5.** Rischi "da intelligenza artificiale" e responsabilità "da reato" dell'ente (d.lgs. 231/2001). – **5.1.** Il "catalogo" dei reati presupposto: le modifiche dirette. – **5.2.** Il "catalogo" dei reati presupposto: le modifiche indirette. – **5.3.** Linee di intervento per la mitigazione del rischio reato "da intelligenza artificiale".

1. Premessa. Dall'AI Act alla normativa nazionale di adeguamento.

L'estate appena trascorsa ha segnato una tappa epocale nel percorso di implementazione, all'interno dell'ordinamento italiano, della normativa in materia di intelligenza artificiale.

In data 10 giugno 2026 il Consiglio dei ministri ha annunciato l'approvazione in via preliminare¹ di un "pacchetto" di due decreti legislativi finalizzati all'adeguamento della normativa nazionale alle disposizioni del regolamento (UE) 2024/1689 (di seguito, "AI Act") nell'esercizio delle deleghe conferite con

* Le opinioni sono espresse dall'Autore a titolo strettamente personale e non vincolano l'Ente di appartenenza.

¹ Comunicato stampa del Consiglio dei Ministri n. 177, in *governo.it*.

la l. 132/2025, recante «*Disposizioni e deleghe al Governo in materia di intelligenza artificiale*» (di seguito, anche "legge AI").

La l. 132/2025 – che, com'è noto, ha rappresentato il primo *step* normativo preordinato all'adeguamento e al coordinamento dell'ordinamento interno alle previsioni dell'AI Act – aveva posto una serie di principi in materia di «*ricerca, sperimentazione, sviluppo, adozione e applicazione di sistemi e di modelli di intelligenza artificiale*», promuovendo «*un utilizzo corretto, trasparente e responsabile, in una dimensione antropocentrica, dell'intelligenza artificiale*» (articolo 1) e conferendo al contempo una serie di deleghe al Governo per l'adeguamento dell'ordinamento interno all'AI Act (articolo 24).

Per quanto di interesse in questa sede, la delega legislativa aveva a oggetto la «*previsione di un'apposita disciplina per l'utilizzo di sistemi di intelligenza artificiale per l'attività di polizia*» (articolo 24, comma 2, lett. h)), l'«*adeguamento del quadro sanzionatorio*» (articolo 24, comma 2, lett. n)), nonché l'adozione di «*uno o più decreti legislativi per adeguare e specificare la disciplina dei casi di realizzazione e di impiego illeciti di sistemi di intelligenza artificiale*».

Sotto quest'ultimo profilo, in particolare, i principi e criteri direttivi al cui esercizio era condizionato l'esercizio della delega erano specificati nella:

- «*introduzione di autonome fattispecie di reato, punite a titolo di dolo o di colpa, incentrate sull'omessa adozione o sull'omesso adeguamento di misure di sicurezza per la produzione, la messa in circolazione e l'utilizzo professionale di sistemi di intelligenza artificiale, quando da tali omissioni deriva pericolo concreto per la vita o l'incolumità pubblica o individuale o per la sicurezza dello Stato*» (articolo 24, comma 5, lett. b));
- «*precisazione dei criteri di imputazione della responsabilità penale delle persone fisiche e amministrativa degli enti per gli illeciti inerenti a sistemi di intelligenza artificiale, che tenga conto del livello effettivo di controllo dei sistemi predetti da parte dell'agente*» (articolo 24, comma 5, lett. c));
- «*nei casi di responsabilità civile, previsione di strumenti di tutela del danneggiato, anche attraverso una specifica regolamentazione dei criteri di ripartizione dell'onere della prova, tenuto conto della classificazione dei sistemi di intelligenza artificiale e dei relativi obblighi come individuati dal regolamento (UE) 2024/1689*» (articolo 24, comma 5, lett. d)).

Si rammenta, infine, come la l. 132/2025 contenesse anche una serie di norme in materia penale direttamente e immediatamente "dispositive", tra le quali (articolo 26):

- una nuova circostanza aggravante comune (articolo 61, comma 1, n. 11-*decies* c.p.), relativa all'aver «*commesso il fatto mediante l'impiego*

di sistemi di intelligenza artificiale, quando gli stessi, per la loro natura o per le modalità di utilizzo, abbiano costituito mezzo insidioso, ovvero quando il loro impiego abbia comunque ostacolato la pubblica o la privata difesa, ovvero aggravato le conseguenze del reato», articolata su inedite forme di impiego di "mezzo fraudolento" e di "minorata difesa" da uso di sistema di intelligenza artificiale;

- di una serie di circostanze aggravanti speciali, riferite, tra gli altri, ai delitti di aggrataggio societario (articolo 2637, comma 2 c.c.) e di manipolazione del mercato (articolo 185, comma 1, secondo periodo d.lgs. 58/1998), per il caso in cui *«il fatto è commesso mediante l'impiego di sistemi di intelligenza artificiale»;*
- di una nuova fattispecie di reato, rappresentata dal delitto di *«Illecita diffusione di contenuti generati o alterati con sistemi di intelligenza artificiale»* (articolo 612-quater c.p.).

I due schemi di decreti legislativi, a cavallo della pausa estiva, sono stati successivamente sottoposti alle Commissioni parlamentari per l'emissione dei pareri di relativa competenza² e, all'esito delle audizioni di un nutrito gruppo di soggetti istituzionali (tra cui AgID, ACN e Garante per la protezione dei dati personali), sono stati approvati in via definitiva dal Consiglio dei ministri in data 4 agosto 2025³.

I provvedimenti sono costituiti dal:

- d.lgs. 9 settembre 2026, n. 160, pubblicato in Gazzetta Ufficiale n. 214 del 15 settembre 2026, recante *«Adeguamento della normativa nazionale alle disposizioni del regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale, in materia di utilizzo dei sistemi di intelligenza artificiale per l'attività di polizia e di responsabilità civile e penale»*, in vigore dal 30 settembre 2026;
- d.lgs. 7 ottobre 2026, n. 179, pubblicato in Gazzetta Ufficiale n. 234 dell'8 ottobre 2026, recante *«Adeguamento della normativa nazionale alle disposizioni del regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale, in materia di poteri delle Autorità nazionali e di utilizzo dell'intelligenza artificiale nella formazione»*, in vigore dal 23 ottobre 2026.

Il "pacchetto" attuativo – secondo le indicazioni rese dal Governo in sede di approvazione preliminare degli schemi di decreti legislativi – si propone come *«una risposta normativa organica alla trasformazione tecnologica in corso»*, che si muove nella direzione di *«promuovere l'innovazione, ma governarla dentro una cornice antropocentrica»*: invero, *«l'IA può sostenere*

² Atto del Governo sottoposto a parere parlamentare n. 421, in [senato.it](#), e Atto del Governo sottoposto a parere parlamentare n. 418, [ivi](#).

³ Comunicato stampa del Consiglio dei Ministri n. 185, in [governo.it](#).

decisioni, servizi, formazione e competitività, ma non sostituire la responsabilità umana né comprimere i diritti fondamentali».

Se «*linea comune delle misure è tenere insieme crescita e garanzie*», anche con il corredo di «*presidi penali per le violazioni più gravi*», «*tratto qualificante dei decreti delegati [...] è l'impostazione antropocentrica*», che si esplica nella costruzione di «*una cornice di garanzie affinché l'innovazione tecnologica resti sempre al servizio della persona, della sua dignità e dei suoi diritti fondamentali*».

Nel merito dei provvedimenti, il d.lgs. 179/2026 disciplina la parte "generale" della normativa attuativa della l. 132/2025, delineando il sistema "istituzionale" delle autorità (nazionali, di vigilanza sul mercato e settoriali) competenti in materia di intelligenza artificiale, le relative prerogative e i connessi poteri (anche di vigilanza e ispettivi), nonché i procedimenti di rispettiva competenza.

Il provvedimento, inoltre – per quanto di interesse ai fini del presente contributo – contiene la disciplina in tema di «*Disposizioni sanzionatorie*» (Sezione V), articolata su misure sanzionatorie non pecuniarie e sanzioni amministrative pecuniarie, rispetto alle quali si profila all'orizzonte un tema di potenziale interferenza (concorso di norme e profili di *bis in idem* sostanziale) con la materia penale.

Il d.lgs. 160/2026 pone una disciplina di natura "speciale" e settoriale relativa, da un lato, all'utilizzo – assai dibattuto in sede di approvazione definitiva del provvedimento – dei sistemi di intelligenza artificiale per l'attività di polizia e, dall'altro, alle disposizioni sanzionatorie in materia penale sostanziale e di responsabilità "da reato" degli enti ai sensi del d.lgs. 231/2001, nonché altre disposizioni in materia processuale penale e civile⁴.

Il presente contributo si propone di fornire un'analisi "a prima lettura" – con riserva di una successiva riflessione più approfondita – delle novità contenute nel d.lgs. 160/2026 (trattando solo incidentalmente la disciplina del d.lgs. 179/2026) con riferimento alle disposizioni di diritto penale sostanziale e in materia di responsabilità "da reato" degli enti ai sensi del d.lgs. 231/2001, formulando alcuni preliminari rilievi e osservazioni rispetto al testo dell'articolato.

2. L'AI Act: impostazione metodologica e ambito di applicazione.

Per comprendere il significato e la portata delle recenti disposizioni di adeguamento dell'ordinamento interno all'AI Act – il quale, com'è noto, costituisce il primo quadro normativo trasversale in materia di intelligenza artificiale adottato a livello sovranazionale – occorre preliminarmente

⁴ Per una prima analisi delle disposizioni penali contenute nello schema del d.lgs. 160/2026, si veda F. RINALDINI, *Intelligenza artificiale: in arrivo nuovi reati e nuove responsabilità per gli enti*, in questa Rivista, 2026, 9.

esaminare, nei limiti della sintesi richiesta dal presente contributo, la struttura e le logiche regolatorie del regolamento europeo.

A livello di "metodo", l'AI Act si propone di istituire a livello europeo un quadro giuridico uniforme per lo sviluppo, l'immissione sul mercato, la messa in servizio e l'uso di sistemi di intelligenza artificiale, promuovendo la diffusione dell'innovazione tecnologica nel mercato interno e garantendo un elevato livello di tutela di beni giuridici di rilevanza primaria, quali la salute, la sicurezza e i diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea (CDFUE), secondo un approccio dichiaratamente "antropocentrico" e fondato sul principio di proporzionalità del rischio (articolo 1 e considerando n. 1).

La scelta del legislatore europeo di introdurre una regolazione di un comparto – quale quello dell'intelligenza artificiale – riconoscendone piena legittimità entro una determinata area di "rischio consentito" implica, com'è naturale, l'allocatione di obblighi di condotta e di divieti in capo ai soggetti della filiera di settore e, correlativamente, un articolato quadro di responsabilità.

L'individuazione del perimetro e del riparto delle responsabilità, nell'ambito della disciplina europea e nazionale – ivi inclusa quella attuativa contenuta nel d.lgs. 160/2026 in commento – presuppone una cursoria analisi della disciplina sostanziale, con specifico riferimento all'ambito di applicazione (oggettivo e soggettivo) dell'AI Act.

Tale analisi – nella chiave di lettura del presente contributo – risulta funzionale alla delimitazione, da un lato, dell'oggetto materiale delle condotte che integrano gli illeciti sanzionatori (in senso lato, nelle varie declinazioni: extrapenale; formalmente penale; sostanzialmente penale) e, dall'altro, secondo una direttrice verticale, delle figure di soggetti "garanti" (nell'accezione estesa, non limitata alla forma del rimprovero omissivo improprio scaturente dall'articolo 40, comma 2 c.p., adottata anche dalla più recente giurisprudenza penale) chiamati a rispondere del fatto illecito.

2.1. Ambito di applicazione oggettivo: la logica risk-based dei sistemi di intelligenza artificiale.

Sotto il profilo oggettivo, l'AI Act (articolo 3, par. 1, nn. 1 e 63) distingue tra «sistema di IA» e «modello di IA per finalità generali» (*general-purpose AI models*), che trovano il proprio *trait d'union* nella nozione di «sistema di IA per finalità generali» (articolo 3, par. 1, n. 66).

Il primo concetto (sistema di IA) è definito dall'AI Act come «un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali».

La nozione di sistema di IA è stata ulteriormente specificata dalle "linee guida" europee sulla «definizione di sistema di intelligenza artificiale»⁵, le quali hanno attuato uno "spacchettamento" di tale nozione in «sette elementi principali» individuabili in:

- 1) automatizzazione («sistema automatizzato»), ossia un sistema che implica lo sviluppo e l'utilizzo di macchine, comprensive delle loro componenti *hardware* e *software*, per eseguire attività quali l'elaborazione dei dati, l'addestramento dei modelli, la modellizzazione predittiva (*predictive modelling*) e il processo decisionale automatizzato su larga scala. L'elemento rilevante non è la tecnologia specifica utilizzata, bensì il fatto che il funzionamento del sistema dipende da capacità di calcolo e di automazione;
- 2) autonomia («progettato per funzionare con livelli di autonomia variabili»), nel senso che il sistema è concepito – come precisato dal considerando n. 12 – per operare con «un certo grado di autonomia di azione rispetto al coinvolgimento umano», nonché di «capacità di funzionare senza l'intervento umano». Ciò implica che, in termini di autonomia, i sistemi di intelligenza artificiale possono spaziare da sistemi che richiedono la supervisione umana a sistemi capaci di svolgere determinate funzioni con un coinvolgimento umano minimo o nullo, con esclusione, per conseguenza, dei sistemi progettati per funzionare esclusivamente con il pieno coinvolgimento e intervento umano manuale;
- 3) adattabilità («che può presentare adattabilità dopo la diffusione»), vale a dire che il sistema può essere dotato di capacità di autoapprendimento, tali da consentirgli di modificare il proprio comportamento nel tempo durante l'utilizzo. Si noti, tuttavia, che questa caratteristica è facoltativa («può»), giacché un sistema può qualificarsi come sistema di intelligenza artificiale anche se, una volta in uso, non continua ad «apprendere autonomamente, scoprire nuovi modelli o individuare relazioni nei dati al di là di quanto inizialmente addestrato a fare»;
- 4) funzionamento «per obiettivi», ossia per «obiettivi dei compiti da svolgere e i loro risultati». Gli obiettivi possono essere «espliciti», quando sono dichiarati e direttamente codificati nel sistema dallo sviluppatore oppure «impliciti», quando non sono esplicitamente indicati, ma possono essere dedotti dal comportamento o dalle ipotesi sottostanti del sistema (ad esempio, possono derivare dai dati di addestramento o dall'interazione del sistema con il proprio ambiente). Si noti, in proposito, che gli obiettivi del sistema possono

⁵ Orientamenti della Commissione sulla definizione di sistema di intelligenza artificiale stabilita dal regolamento (UE) 2024/1689 (regolamento sull'IA), 29 luglio 2025.

- essere diversi dalla sua «*finalità*», la quale ha una valenza esterna e comprende il contesto in cui il sistema è progettato per essere impiegato e le sue modalità di funzionamento;
- 5) capacità inferenziale («*deduce dall'input che riceve come generare l'output*»), intesa – nell'accezione sempre espressa dal considerando n. 12, in linea, tra l'altro, con la norma tecnica IOS/IEC 22989:2022 (*Information technology – Artificial intelligence – Artificial intelligence concepts and terminology*) – come la capacità di determinare come produrre un risultato sulla base degli *input* ricevuti. Tale caratteristica distingue i sistemi di intelligenza artificiale dai *software* “tradizionali” o dagli approcci di programmazione più semplici, che si limitano ad applicare regole fisse definite esclusivamente da persone fisiche;
 - 6) capacità di generare *output* e tipologia di *output*, che ricomprendono:
 - a. previsioni, che costituiscono la stima di un valore sconosciuto (*output*) a partire da valori noti forniti al sistema (*input*);
 - b. contenuti, che si riferiscono alla generazione di nuovo materiale da parte del sistema (ad esempio, testi, immagini, audio, video o altri materiali);
 - c. raccomandazioni, che sono rappresentate da suggerimenti su azioni, prodotti o servizi rivolti agli utenti sulla base delle loro preferenze, comportamenti o di altri *input* di dati;
 - d. decisioni, che sono da intendersi come conclusioni o scelte effettuate da un sistema.
 - 7) interazioni con l'ambiente («*possono influenzare ambienti fisici o virtuali*»), nel senso che il sistema non è passivo, ma ha un impatto attivo sull'ambiente in cui è impiegato. Tale influenza può manifestarsi in ambienti fisici (ad esempio, mediante il controllo di macchinari, *robot* o dispositivi) oppure in ambienti virtuali (come piattaforme digitali, sistemi informatici, flussi di dati o ecosistemi *software*).

Il secondo concetto (modello di IA per finalità generali) fa riferimento a modelli di intelligenza artificiale addestrati “su larga scala” (cioè con grandi quantità di dati) e destinati a svolgere una pluralità di compiti distinti, i quali possono essere integrati in una varietà di sistemi di intelligenza artificiale o di applicazioni a valle. Rientrano in tale categoria, per notorietà, i moderni modelli linguistici di grandi dimensioni (*large language model*, LLM).

Premesse tali necessarie categorizzazioni preliminari, si osserva come caratteristica qualificante dell'AI Act sia l'adozione di un modello di regolazione *risk based*, nel quale l'intensità degli obblighi normativi e dei controlli portati dagli stessi aumenta progressivamente in funzione del rischio che il sistema di intelligenza artificiale presenta per i diritti fondamentali (sanciti dalla CDFUE).

In tale prospettiva, possono distinguersi i sistemi di intelligenza artificiale in relazione a quattro principali livelli di rischio (profilo qualitativo, sebbene modulato sulla base di un criterio quantitativo).

Al livello più elevato si collocano i sistemi di intelligenza artificiale che integrano un rischio “inaccettabile” (indicati come «*pratiche di AI vietate*» dall’articolo 5 AI Act), per i quali il regolamento europeo prevede un divieto pressoché assoluto di immissione sul mercato, di messa in servizio o di utilizzo⁶.

Rientrano in tale categoria, tra gli altri, i sistemi di manipolazione comportamentale che utilizzano «*tecniche subliminali, manipolative o ingannevoli*» (lett. a), quelli che sfruttano le «*vulnerabilità di una persona fisica o di uno specifico gruppo di persone*» connesse all’età, alla disabilità o alla condizione sociale o economica (lett. b), talune forme di *social scoring*, inteso come «*classificazione delle persone fisiche o gruppi di persone per un determinato periodo di tempo sulla base del loro comportamento sociale o di caratteristiche personali o della personalità*» (lett. c), specifiche applicazioni di “polizia predittiva” (ossia finalizzate a «*valutare o prevedere il rischio che una persona fisica commetta un reato*») basate esclusivamente sulla profilazione (lett. d) e numerose pratiche di identificazione o di categorizzazione biometrica (*facial scraping*) ritenute incompatibili con i diritti fondamentali. Un secondo livello è rappresentato dai sistemi «*ad alto rischio*» (articoli 6 ss. AI Act) – che costituiscono il vero fulcro, anche dal punto di vista “quantitativo”, della disciplina regolatoria europea e, come si vedrà oltre, di quella attuativa nazionale – nell’ambito dei quali rientrano:

- sia i sistemi destinati a essere utilizzati come componenti di sicurezza di un prodotto o costituenti essi stessi un prodotto, già soggetto alla normativa europea di armonizzazione dell’Allegato I (articolo 6, par. 1, lett. a) AI Act);
- sia il prodotto il cui componente di sicurezza è il sistema di intelligenza artificiale o il sistema di intelligenza artificiale stesso, in quanto prodotto, già soggetto alla normativa europea di armonizzazione dell’Allegato I (articolo 6, par. 1, lett. b) AI Act);
- sia quelli (articolo 6, par. 2 AI Act), richiamati nell’Allegato III, impiegati in settori particolarmente “sensibili”, quali, tra gli altri, infrastrutture critiche (digitali, traffico stradale, fornitura di acqua, gas, riscaldamento o elettricità), istruzione e formazione, occupazione, gestione dei rapporti di lavoro, accesso a servizi privati essenziali e a prestazioni e servizi pubblici essenziali e fruizione degli stessi (ad

⁶ Le disposizioni relative alle «*pratiche di IA vietate*» (Capo II AI Act), ai sensi dell’articolo 113, par. 1, lett. a) AI Act – come di recente modificato dal Regolamento (UE) 2026/1744 (*Digital Omnibus on AI*) – sono applicabili a partire dal 2 febbraio 2025 (ad eccezione di alcune specifiche disposizioni applicabili a far data dal prossimo 2 dicembre 2026).

esempio, sistemi destinati a essere utilizzati per valutare l'affidabilità creditizia delle persone fisiche o per stabilire il loro merito di credito), gestione delle migrazioni e amministrazione della giustizia.

Si evidenzia che, allo stato attuale, non è stata ancora adottata, all'esito di consultazione pubblica, la versione definitiva delle "linee guida" europee sulla classificazione ad alto rischio dei sistemi di intelligenza artificiale⁷, le quali – una volta approvate – costituiranno un utilissimo strumento interpretativo non soltanto ai fini dell'analisi regolatoria, ma anche, come oltre si illustrerà, della comprensione del perimetro applicativo delle fattispecie penali di nuova introduzione da parte del d.lgs. 160/2026.

Per tali sistemi il legislatore europeo impone un articolato apparato di *compliance*, che comprende un eterogeneo sistema di presidi costituiti da:

- sistemi di gestione del rischio (articolo 9 AI Act);
- requisiti di qualità e di *governance* dei dati (articolo 10 AI Act);
- documentazione tecnica (articolo 11 AI Act);
- registrazione automatica degli eventi (*logging*, articolo 12 AI Act);
- obblighi di trasparenza (tecnica) verso gli utilizzatori (articolo 13 AI Act);
- «*sorveglianza umana*» (articolo 14 AI Act);
- requisiti di accuratezza, di robustezza e di cybersicurezza (articolo 15 AI Act);
- sistema di gestione della qualità (articolo 17 AI Act);
- conservazione dei documenti (articolo 18 AI Act);
- valutazione di impatto sui diritti fondamentali (*fundamental right impact assessment*, FRIA, articolo 27 AI Act);
- dichiarazione di conformità UE (articolo 47 AI Act);
- marcatura CE (articolo 48 AI Act);
- registrazione (articoli 49 e 71 AI Act);
- procedure di valutazione della conformità (c.d. *conformity assessment*, articolo 43 AI Act);
- monitoraggio successivo all'immissione sul mercato (*post market monitoring*, articolo 72 AI Act).

Le misure di «*sorveglianza umana*» (articolo 14 AI Act), in particolare, fanno riferimento alla previsione di presidi, in fase di progettazione e di sviluppo del sistema di intelligenza artificiale, affinché gli stessi possano essere «*efficacemente supervisionati da persone fisiche durante il periodo in cui sono in uso*», se del caso «*anche con strumenti di interfaccia uomo-macchina adeguati*».

La finalità del requisito (c.d. *human oversight*) è «*prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali che possono emergere quando un sistema di IA ad alto rischio è utilizzato*», in via alternativa,

⁷ Draft Commission guidelines on the classification of high-risk AI systems under Article 6 of Regulation (EU) 2024/1689 (AI Act) for stakeholder consultation, 2026.

«conformemente alla sua finalità prevista» (uso "proprio") o «in condizioni di uso improprio ragionevolmente prevedibile» (uso "improprio" prevedibile).

Le misure di sorveglianza umana, peraltro, si applicano nel rispetto di un generale canone di sussidiarietà, nel senso che sono doverose *«in particolare qualora tali rischi persistano nonostante l'applicazione di altri requisiti»* previsti per i sistemi ad alto rischio.

Esse, inoltre, devono trovare implementazione secondo un criterio di proporzionalità, giacché *«sono commisurate ai rischi, al livello di autonomia e al contesto di utilizzo del sistema»*.

Quanto alle modalità di loro attuazione, esse sono garantite, in via disgiunta, attraverso l'adozione di:

- *«misure individuate e integrate nel sistema di IA ad alto rischio dal fornitore prima della sua immissione sul mercato o messa in servizio», «ove tecnicamente possibile»* (sorveglianza umana "a monte");
- *«misure individuate dal fornitore prima dell'immissione sul mercato o della messa in servizio del sistema di IA ad alto rischio, adatte ad essere attuate dal deployer»* (sorveglianza umana "a valle" o "derivata").

Si noti, peraltro – punto assai rilevante anche ai fini delle disposizioni penali che saranno oltre esaminate – che l'applicabilità delle disposizioni dell'AI Act relative ai sistemi di intelligenza artificiale ad alto rischio (classificazione, requisiti, obblighi dei fornitori, dei *deployer* e di altre parti), originariamente prevista a partire dallo scorso 2 agosto 2026 (articolo 113, par. 1 AI Act, con l'eccezione già prevista dalla rispettiva lett. c) per i sistemi di cui all'Allegato I, per le cui disposizioni era prevista l'applicabilità a partire dal 2 agosto 2027), è stata recentemente posticipata, per l'effetto delle modifiche apportate dal Regolamento (UE) 2026/1744 (*Digital Omnibus on AI*), secondo l'attuale previsione dell'articolo 113, par. 1, lett. c) AI Act:

- al 2 dicembre 2027 per quanto riguarda i sistemi di cui all'Allegato III;
- al 2 agosto 2028 per quanto riguarda i sistemi di cui all'Allegato I.

Due ulteriori livelli – benché non espressamente codificati dall'AI Act – compongono la "scala del rischio" dei sistemi di intelligenza artificiale

Appartengono a una terza categoria i sistemi di intelligenza artificiale a rischio "limitato" (o di trasparenza), assoggettati prevalentemente – per l'appunto – a obblighi di trasparenza (intesa come trasparenza "informativa"). Rientrano in tale ambito, ad esempio, i sistemi che interagiscono direttamente con gli utenti (ad esempio, i *chatbot*), quelli capaci di generare contenuti sintetici (*AI generated content*) o di manipolare immagini, audio e video (*deepfake*).

In tali ipotesi, l'utente finale (*final user*, come denominato nelle versioni non definitive dell'AI Act) deve essere informato dell'interazione con un sistema di intelligenza artificiale o della natura artificiale dei contenuti, così da garantirgli la piena consapevolezza e da prevenire fenomeni di inganno o di manipolazione informativa.

Alla base dell'ideale "piramide dei rischi", infine, si collocano i sistemi "a rischio minimo" o "trascurabile" (quali, ad esempio, i filtri *spam*), per i quali l'AI Act non introduce obblighi cogenti, ulteriori rispetto alle normative già applicabili (tra le altre, il GDPR).

2.2. Ambito di applicazione soggettivo: la «catena del valore» dell'intelligenza artificiale.

Una volta accertata la riconducibilità di un applicativo alla definizione di sistema di intelligenza artificiale (profilo oggettivo) e identificato il livello di rischio associato (profilo qualitativo), secondo la logica dell'AI Act, il concreto riparto delle responsabilità lungo la filiera dell'intelligenza artificiale impone l'analisi del profilo soggettivo, ossia l'individuazione delle figure "tipiche" titolari di specifici obblighi giuridici (e divieti) nelle varie fasi di "vita" dei sistemi di intelligenza artificiale, i quali finiscono per conformare dall'esterno le rispettive "sfere" di gestione del rischio, rilevanti nella prospettiva di allocazione delle varie forme di responsabilità (civilistiche e sanzionatorie in senso lato) che possono scaturire per l'ipotesi di violazione dei suddetti obblighi.

Dal punto di vista soggettivo, l'architettura dell'AI Act si fonda sulla distinzione tra diverse categorie di «operatori» (categoria onnicomprensiva definita nell'articolo 3, par. 1, n. 8 AI Act) – che costituiscono la «catena del valore» dell'intelligenza artificiale (considerando n. 9, 20, 83, 85, 88, 89, 90, 101, 109, 110, 114, 135) – ai quali sono attribuiti obblighi differenziati in ragione della posizione rivestita e del ruolo concretamente svolto, nonché dei poteri di intervento rispetto al sistema di intelligenza artificiale.

La filiera dell'intelligenza artificiale prevede, anzitutto, a un livello idealmente sovraordinato, due figure "chiave", per rilevanza e per responsabilità.

Centrale è la figura del «fornitore», altresì detto *provider* (articolo 3, par. 1, n. 3 AI Act), ossia la persona fisica o giuridica, l'autorità pubblica, l'agenzia o un altro organismo che si qualifica per l'esercizio delle seguenti prerogative, da intendersi cumulative:

- «sviluppa o fa sviluppare» un sistema di intelligenza artificiale o un modello di intelligenza artificiale per finalità generali;
- lo «immette [...] sul mercato» o lo «mette in servizio», in ogni caso «con il proprio nome o marchio, a titolo oneroso o gratuito».

Altra figura prioritaria, lungo la catena dell'intelligenza artificiale, accanto al *provider*, è il «deployer» (articolo 3, par. 1, n. 4 AI Act), identificabile nella persona fisica o giuridica, nell'autorità pubblica, nell'agenzia o in altro organismo che «utilizza» il sistema di intelligenza artificiale «sotto la propria autorità» e nel contesto della propria attività professionale (escludendo giustappunto dalla soggettività giuridica in esame la casistica in cui il sistema di intelligenza artificiale «sia utilizzato nel corso di un'attività personale non professionale»).

Parimenti rilevante, nella prospettiva di gestione dei rischi da intelligenza artificiale, è il «fornitore a valle» (c.d. *downstream provider*), ossia il «fornitore di un sistema di IA, compreso un sistema di IA per finalità generali, che integra un modello di IA», tra l'altro «indipendentemente dal fatto che il modello di IA sia fornito dallo stesso e integrato verticalmente o fornito da un'altra entità sulla base di relazioni contrattuali» (articolo 3, par. 1, n. 68 AI Act).

L'AI Act, poi, identifica ulteriori figure, che si pongono in una posizione idealmente collaterale – ma per ciò non meno rilevante per la filiera di settore – rispetto al *provider* e al *deployer*, a carico dei quali sono previsti specifici obblighi di cooperazione, di controllo e di tracciabilità:

- l'«importatore», ossia la persona fisica o giuridica «ubicata o stabilita nell'Unione europea» che «immette sul mercato» un sistema di intelligenza artificiale «recante il nome o il marchio di una persona fisica o giuridica stabilita in un paese terzo» (articolo 3, par. 1, n. 6 AI Act);
- il «distributore», ossia la persona fisica o giuridica nella catena di approvvigionamento, «diversa dal fornitore o dall'importatore», che «mette a disposizione [...] sul mercato dell'Unione» un sistema di intelligenza artificiale (articolo 3, par. 1, n. 7 AI Act);
- il «rappresentante autorizzato», ossia la persona fisica o giuridica ubicata o stabilita nell'Unione europea, che «ha ricevuto e accettato un mandato scritto da un fornitore» di un sistema di intelligenza artificiale o di un modello di intelligenza artificiale per finalità generali «al fine, rispettivamente, di adempiere ed eseguire per suo conto gli obblighi e le procedure» stabiliti dall'AI Act (articolo 3, par. 1, n. 5 AI Act).

Si rammenta, inoltre, che l'AI Act pone un'importante «clausola di equivalenza» (articolo 25 AI Act), in forza della quale – nell'ottica di irrobustire il quadro delle responsabilità secondo un principio di «prossimità» al rischio – un soggetto qualificabile come distributore, importatore, *deployer* o altro terzo «è considerato fornitore» di un sistema di intelligenza artificiale (solo) ad alto rischio in presenza di una delle seguenti condizioni:

- a) se «appone il proprio nome o marchio» su un sistema di intelligenza artificiale ad alto rischio già immesso sul mercato o messo in servizio;
- b) se «apporta una modifica sostanziale» a un sistema di intelligenza artificiale ad alto rischio già immesso sul mercato o messo in servizio, intendendosi per tale la modifica di un sistema di intelligenza artificiale «che non è prevista o programmata nella valutazione iniziale della conformità effettuata dal fornitore» e che «ha l'effetto di incidere sulla conformità del sistema di IA» rispetto ai requisiti previsti per i sistemi ad alto rischio (Capo III, Sezione 2) o che «comporta una modifica della finalità prevista per la quale il sistema di IA è stato valutato» (articolo 3, par. 1, n. 23 AI Act);

- c) o se modifica la finalità prevista di un sistema di intelligenza artificiale, anche un sistema per finalità generali.

A fronte di tale articolazione soggettiva della filiera dell'intelligenza artificiale, l'AI Act modula gli obblighi giuridici posti a carico degli «operatori» in funzione di una serie combinata di fattori, rappresentati, rispettivamente, dal ruolo assunto nella catena del valore, dalla tipologia di "tecnologia" impiegata (sistema di intelligenza artificiale o modello di intelligenza artificiale per finalità generali) e dal livello di rischio associato al sistema di intelligenza artificiale.

Tale meccanismo di graduazione "integrata" della disciplina regolatoria implica che un medesimo soggetto – in una logica di "simmetria variabile" – può essere destinatario di obblighi (e di responsabilità) differenti a seconda che sviluppi o utilizzi un sistema di intelligenza artificiale ad alto rischio, un sistema di intelligenza artificiale a rischio "limitato" soggetto a meri obblighi di trasparenza ovvero un modello di intelligenza artificiale per finalità generali.

Nella dinamica regolatoria dell'AI Act, il fornitore rappresenta il principale centro di imputazione degli obblighi e delle conseguenti responsabilità connessi alla *compliance* con la normativa in materia di intelligenza artificiale. Con particolare riferimento ai sistemi di intelligenza artificiale ad alto rischio – di diretto interesse ai fini della successiva analisi delle disposizioni in materia penale – esso è tenuto ad adempiere gli obblighi previsti dall'articolo 16 AI Act, quali, tra gli altri:

- implementare un «sistema di gestione del rischio», inteso come «un processo iterativo continuo pianificato ed eseguito nel corso dell'intero ciclo di vita» del sistema (articolo 9 AI Act);
- garantire «pratiche di governance e gestione dei dati», utilizzati ai fini di addestramento, convalida e prova, che siano «adeguate» alla finalità prevista del sistema (articolo 10 AI Act);
- predisporre e mantenere aggiornata la documentazione tecnica del sistema (articolo 11 AI Act);
- progettare il sistema in modo da consentire un'effettiva «supervisione umana» (articolo 14 AI Act);
- garantire un «adeguato livello di accuratezza, robustezza e ciphersicurezza» del sistema (articolo 15 AI Act);
- svolgere le «procedure di valutazione della conformità» rispetto ai requisiti previsti per i sistemi di intelligenza artificiale ad alto rischio (articolo 16, par. 1, lett. f) AI Act, in relazione all'articolo 43 AI Act);
- procedere alla «marcatura CE» (articolo 16, par. 1, lett. h) AI Act, in relazione articolo 48 AI Act) e alla «registrazione» del sistema nella banca dati dell'Unione europea (articolo 16, par. 1, lett. i) AI Act, in relazione all'articolo 49 AI Act).

Si noti, inoltre, che gli obblighi di conformità del fornitore non si esauriscono nella fase antecedente all'immissione sul mercato del sistema di intelligenza artificiale (*ex ante*), ma si estendono alla fase successiva (*ex post*).

In particolare, nella fase *post-market*, è previsto a carico del fornitore l'obbligo di istituire un «*sistemi di monitoraggio*» (articolo 72 AI Act), finalizzato a raccogliere e analizzare le informazioni relative al funzionamento e alle prestazioni del sistema, nonché un obbligo di «*comunicazione*» verso le autorità di vigilanza del mercato, in forza del quale i fornitori «*segnalano qualsiasi incidente grave*», ossia un c.d. *AI breach* (articolo 73 AI Act).

La figura del *deployer* – nella quale, a livello pratico, saranno generalmente sussumibili le imprese utilizzatrici di sistemi di intelligenza artificiale – è destinataria, parimenti, di rilevanti obblighi di *compliance* alla normativa dell'AI Act.

In forza della norma generale di riferimento (articolo 26 AI Act), riferita ai sistemi di intelligenza artificiale ad alto rischio, il *deployer* è tenuto, anzitutto, a «*utilizzare tali sistemi conformemente alle istruzioni per l'uso*» definite dal fornitore, adottando «*idonee misure tecniche e organizzative*».

Egli, inoltre, deve a garantire che le persone incaricate dell'utilizzo del sistema dispongano di adeguata competenza e formazione, monitorare il funzionamento dello stesso e, qualora individui un incidente grave, deve informare, in osservanza di un onere di collaborazione il fornitore (nonché l'importatore, il distributore e le autorità di vigilanza del mercato).

Di particolare significatività è la previsione (articolo 27 AI Act), per cui il *deployer* – nel rispetto delle specifiche condizioni ivi descritte, sempre riferibili ai soli sistemi di intelligenza artificiale ad alto rischio (tra cui quelli finalizzati, ad esempio, alla valutazione del merito creditizio e del rischio assicurativo) – è tenuto a effettuare la valutazione di impatto sui diritti fondamentali (FRIA), ossia un *assessment* preordinato ad analizzare le ricadute dell'utilizzo del sistema di intelligenza artificiale sui diritti fondamentali degli utenti che con gli stessi sono destinati a interagire (*affected persons*) e a definire le conseguenti misure di mitigazione.

Gli importatori e i distributori, d'altro canto – che non intervengono nella fase di sviluppo del sistema di intelligenza artificiale – sono a propria volta chiamati a svolgere importanti funzioni di presidio della conformità, con particolare riferimento ai sistemi di intelligenza artificiale ad alto rischio, nell'ambito di un ideale sistema di controlli, rispettivamente, “di secondo” e di “terzo” livello (articoli 23 e 24 AI Act).

Essi, per la parte di propria competenza, devono verificare che il fornitore abbia eseguito la procedura di valutazione della conformità, che il sistema di intelligenza artificiale sia provvisto della richiesta documentazione tecnica, che lo stesso rechi la marcatura CE e sia accompagnato dalla dichiarazione di conformità UE e – quale obbligo di “chiusura” a carico del distributore – che il fornitore e l'importatore abbiano ottemperato i rispettivi obblighi.

Un complesso specifico di obblighi, infine, riguarda – trasversalmente, senza distinzione per il livello di rischio – «*determinati*» sistemi di intelligenza artificiale, individuati nell’articolo 50 AI Act, rappresentati da:

- sistemi «*destinati a interagire direttamente con le persone fisiche*», sistemi «*che generano contenuti audio, immagine, video o testuali sintetici*»;
- sistemi «*di riconoscimento delle emozioni o di un sistema di categorizzazione biometrica*»;
- un sistema «*che genera o manipola immagini o contenuti audio o video che costituiscono un «deep fake»*»;
- un sistema «*che genera o manipola testo pubblicato allo scopo di informare il pubblico su questioni di interesse pubblico*».

Rispetto a tale categoria è previsto a carico – a seconda dei casi – del fornitore o del *deployer* un obbligo di informazione o di marcatura (*labelling*) in favore delle persone fisiche che interagiscono direttamente con tali sistemi, al fine di renderle edotte dell’interazione con il sistema medesimo o che l’*output* è stato generato o manipolato artificialmente.

Il *framework* normativo sinteticamente descritto assume una diversa fisionomia in relazione ai modelli di intelligenza artificiale per finalità generali (GPAI), per i quali il legislatore europeo ha previsto uno statuto parzialmente divaricato rispetto alla disciplina “ordinaria” (articoli 51 ss. AI Act)⁸, nella direzione di introdurre specifici obblighi direttamente in capo al fornitore del modello, differenziati a seconda che si tratti di modelli “a rischio basico” o «*a rischio sistemico*».

Tra gli obblighi posti a carico del fornitore di GPAI “a rischio basico” (articolo 53 AI Act) rilevano, in particolare, la predisposizione e l’aggiornamento della documentazione tecnica del modello, la messa a disposizione delle informazioni necessarie ai fornitori di sistemi di intelligenza artificiale che integrano il modello nei propri sistemi (c.d. *downstream provider*), l’adozione di politiche interne finalizzate al rispetto della normativa sul diritto d’autore e la pubblicazione di una sintesi sufficientemente dettagliata dei contenuti utilizzati per l’addestramento dei modelli.

Gli obblighi del fornitore si intensificano, secondo un *climax* ascendente, nel caso di GPAI qualificati come modelli «*a rischio sistemico*» – ossia quelli che presentano «*un rischio specifico per le capacità di impatto elevato dei modelli di IA per finalità generali, avente un impatto significativo sul mercato dell’Unione a causa della sua portata o di effetti negativi effettivi o ragionevolmente prevedibili sulla salute pubblica, la sicurezza, i diritti*

⁸ Le disposizioni relative ai modelli di intelligenza artificiale per finalità generali (GPAI) (Capo V AI Act), ai sensi dell’articolo 113, par. 1, lett. b) AI Act, sono applicabili a partire dal 2 agosto 2025 (ad eccezione delle relative disposizioni sanzionatorie contenute nell’articolo 101 AI Act, per cui il termine è stato fissato al 2 agosto 2026).

fondamentali o la società nel suo complesso, che può propagarsi su larga scala lungo l'intera catena del valore» (articolo 3, par. 1, n. 65 AI Act), in costanza delle condizioni indicate nell'articolo 51 AI Act – rispetto ai quali il fornitore è chiamato a porre in essere attività periodiche di valutazione e di *testing* del modello, a individuare e mitigare i rischi sistemici, a monitorare e a comunicare gli incidenti significativi, nonché ad adottare misure rafforzate di *cybersecurity* e di gestione del rischio (articolo 55 AI Act).

3. L'AI Act: *framework* sanzionatorio.

In funzione di "chiusura" del sistema regolatorio sopra illustrato, il legislatore europeo ha inteso assicurare l'effettività del sistema di *governance* dell'intelligenza artificiale ricorrendo a un articolato apparato sanzionatorio (Capo XII AI Act), fondato, in un'ottica di proporzionalità, sulla medesima logica *risk-based* che permea l'architettura oggettiva e soggettiva dell'AI Act. Preme evidenziare, da un punto di vista generale, che il legislatore europeo non ha direttamente definito le sanzioni applicabili per le violazioni dell'AI Act, ma ha demandato agli Stati membri il compito di stabilire le norme «*relative alle sanzioni e alle altre misure di esecuzione, che possono includere anche avvertimenti e misure non pecuniarie*», seppur «*[i]n conformità dei termini e delle condizioni*» stabilite dall'articolo 99 AI Act, il quale, nell'individuare i profili di violazione rilevanti, si limita a definire le soglie edittali massime.

Occorre evidenziare, in proposito, che le sanzioni connesse alla violazione della normativa in materia di intelligenza artificiale, oltre a essere «*effettive, proporzionate e dissuasive*», in una prospettiva di gradualità e di esigibilità in concreto, devono tenere conto «*degli interessi delle PMI, comprese le start-up, e della loro sostenibilità economica*».

3.1. Le misure sanzionatorie amministrative.

Per quanto concerne le sanzioni amministrative pecuniarie, l'AI Act ripropone il metodo di commisurazione della cornice edittale sul modello – oggi diffuso anche in altri provvedimenti normativi di matrice europea e già sperimentato, ad esempio, nel regolamento (UE) 2016/679 (GDPR) – che fonda la sanzione, a seconda dei casi, su un criterio quantitativo-numerico ovvero, se l'autore dell'illecito «*è un'impresa*», ove l'importo finale risultante dal calcolo sia «*superiore*» a quello su base numerica, su un criterio di tipo percentuale, ricavato sul fatturato mondiale annuo.

I massimi delle sanzioni amministrative pecuniarie spaziano, in funzione della gravità della violazione, secondo un criterio decrementale, all'interno dei seguenti *range*:

- fino al 7% del fatturato mondiale annuo (o, comunque, fino a 35 milioni di euro) per le violazioni relative alle pratiche vietate (rischio “inaccettabile”) di cui all’articolo 5 AI Act;
- fino al 3% del fatturato mondiale annuo (o, comunque, fino a 15 milioni di euro) per l’inosservanza degli obblighi relativi ai sistemi ad “alto rischio” o di trasparenza posti, a seconda dei casi, a carico dei fornitori, dei *deployer*, degli importatori, dei distributori e dei rappresentanti autorizzati;
- fino all’1% del fatturato mondiale annuo (o, comunque, fino a 7,5 milioni di euro) per la fornitura di informazioni inesatte, incomplete o fuorvianti agli organismi notificati o alle autorità nazionali competenti per dare seguito a una richiesta.

Con riferimento alle violazioni afferenti ai GPAI (articolo 101 AI Act), i massimi edittali previsti a carico dei fornitori – fino al 3% del fatturato mondiale annuo totale o fino a 15 milioni di euro – sono stabiliti secondo il medesimo metodo alternativo, con la particolarità che, a livello terminologico, la scelta tra criterio numerico e criterio percentuale non è espressamente basata sulla natura di «*impresa*» o meno del soggetto sanzionato (che, essendo un «*fornitore*», ragionevolmente sarà una persona giuridica esercente attività di impresa).

Il descritto quadro sanzionatorio a livello europeo ha trovato specificazione – in linea a quanto previsto dall’articolo 99 AI Act – nelle disposizioni del d.lgs. 179/2026, il quale nella Sezione V, rubricata «*Disposizioni sanzionatorie*», oltre a definire i poteri sanzionatori in capo all’Autorità nazionale di notifica (AgID) e alle Autorità nazionali di vigilanza del mercato (articoli 19, 20 e 21), articola un sistema di «*misure sanzionatorie*», rispettivamente, «*non pecuniarie*» e «*pecuniarie*».

In proposito, l’articolo 22 d.lgs. 179/2026 individua il novero delle misure sanzionatorie «*non pecuniarie*», le quali, in ragione della propria natura (appunto, non pecuniaria), si caratterizzano per un peculiare regime applicativo ispirato ai principi di sussidiarietà e di progressività, volto a valorizzare le caratteristiche di “minimalità” dell’illecito, nell’ambito di un giudizio “discrezionale” dell’autorità competente.

In tal senso, le misure sanzionatorie non pecuniarie «*possono*» – e non devono – trovare applicazione, in primo luogo, in relazione alle «*violazioni del regolamento IA caratterizzate da scarsa offensività o pericolosità*».

Si osserva, sul punto, che, pur a fronte del testo normativo impostato in termini di facoltatività («*possono*»), la previsione di un requisito oggettivo a cui è subordinata l’applicazione delle misure non pecuniarie parrebbe introdurre un canone di discrezionalità comunque “vincolata” in capo all’autorità competente, nel senso di imporre – e non di facoltativizzare – l’applicazione di tale tipologia di misure in caso di violazione di scarsa offensività o pericolosità.

Nondimeno, un'interpretazione alternativa potrebbe ritenere, invece, che, anche a fronte di violazioni di tal guisa, astrattamente l'autorità competente potrebbe optare per l'applicazione di sanzioni pecuniarie "ordinarie", in luogo di quelle non pecuniarie, fermo restando l'esercizio di uno specifico dovere motivazionale sul punto.

Quanto al requisito applicativo delle misure non pecuniarie, il lemma normativo non pare particolarmente felice nella propria formulazione, con riferimento al presupposto applicativo delle stesse (*«scarsa offensività o pericolosità»*), dal momento che l'offesa e il pericolo notoriamente si pongono in rapporto di genere a specie, essendo il pericolo una delle modalità con cui si può esplicare l'offesa, insieme al danno, da intendersi nell'accezione di una lesione soltanto potenziale ai beni giuridici di riferimento.

Tanto premesso, il requisito applicativo in esame allude evidentemente a una violazione che, per caratteristiche oggettive (modalità della condotta, natura e dimensioni del soggetto agente, entità delle conseguenze derivanti dalla violazione, tenuto conto anche di eventuali misure riparatorie e ripristinatorie già adottate, anche in via urgente) e soggettive (gradazione dell'elemento psichico), risulti "minimale" e, per ciò, emendabile, oltre che meritevole – in una prospettiva di giustizia sostanziale – di sanzioni non incidenti (quanto meno in via diretta) sulla sfera economico-patrimoniale del trasgressore.

L'applicazione delle misure non pecuniarie, inoltre, deve avvenire *«in linea con i criteri richiamati all'articolo 23, comma 6»*, ossia – svolgendo il rinvio normativo – secondo i criteri previsti dall'articolo 99, par. 7 Al Act e tenendo in conto la *«natura, sostanziale o formale, dell'obbligo violato»*.

Il riferimento vincolante ai criteri dettati dall'articolo 99, par. 7 Al Act, tuttavia, deve essere correttamente interpretato in relazione alle misure in esame, poiché, mentre per le sanzioni pecuniarie – nel cui ambito di disciplina (articolo 23) il rinvio è contenuto – i criteri sono funzionali alla *«determinazione dell'importo della sanzione amministrativa pecuniaria»* (cioè al *quantum*), per le misure non pecuniarie il richiamo mediato all'Al Act sembra dover essere inteso in relazione all'*an*, ossia come criterio discrezionale al fine di (come previsto dall'articolo 99, par. 7 Al Act) *«decidere se infliggere una sanzione amministrativa pecuniaria»* oppure una misura non pecuniaria. Una lettura alternativa – ma forse meno ragionevole – potrebbe intendere il richiamo ai criteri dell'articolo 99, par. 7 Al Act come finalizzato a introdurre degli "indici" di graduazione, sotto il profilo dell'afflittività – dunque nella stessa logica della disposizione dell'articolo 23 in tema di sanzioni pecuniarie – anche in sede di scelta della misura non pecuniaria in concreto da applicare. Il rinvio a detti criteri potrebbe essere inteso, ancor prima e forse più condivisibilmente, come preordinato all'individuazione di specifici "indici" – non di graduazione della misura – bensì di accertamento stesso del requisito applicativo della *«scarsa offensività o pericolosità»*, il cui apprezzamento, in

effetti, è del tutto sfornito di una base valutativa ancorata a referenti in fatto o in diritto tipizzati normativamente.

Una tesi "mista" o "intermedia", infine, potrebbe indurre a ritenere che i criteri dell'articolo 99, par. 7 Al Act, in relazione alle misure non pecuniarie, svolgano una triplice (o almeno duplice) funzione "scalare": indice di accertamento del requisito della scarsa offensività o pericolosità; criterio dirimente per l'opzione in favore di un trattamento sanzionatorio pecuniario o soltanto non pecuniario; in caso di opzione per il regime non pecuniario, criterio di scelta della misura non pecuniaria concreta (tra quelle astrattamente ipotizzabili). Come detto in precedenza, infine, le misure non pecuniarie sono ispirate a un regime di sussidiarietà del sistema sanzionatorio, essendo destinate a operare *«in alternativa all'applicazione delle sanzioni amministrative pecuniarie»* (di cui al successivo articolo 23) e mai – parrebbe – in maniera combinata.

Nel merito, le misure non pecuniarie possono consistere, da un lato, nell'ordine – con funzione al contempo riparatoria e impeditiva – di *«eliminare le infrazioni e di astenersi dal ripeterle»* (lett. a), il quale, se del caso, potrà anche indicare, in forma prescrittiva specifica, *«le misure da adottare e il termine per attuarle»*; dall'altro, in una misura dalla spiccata portata "reputazionale" – che in concreto potrebbe risultare ben più afflittiva, in ragione delle proprie ricadute dal punto di vista della *brand reputation* della società, non solo dell'ordine ripristinatorio, ma anche delle sanzioni pecuniarie – consistente nella *«dichiarazione pubblica avente ad oggetto la violazione commessa e il soggetto responsabile»* (lett. b), applicabile per le violazioni "a condotta esaurita" (rispetto alle quali nulla vi sia da regolarizzare), ossia *«qualora l'infrazione contestata sia cessata»*.

L'eventuale inosservanza dell'ordine ripristinatorio, poi, è presidiata da una sanzione di secondo livello "derivata", consistente – proprio in ragione del contegno inerte del trasgressore e dell'ineffettività della tutela non pecuniaria – nell'applicazione delle sanzioni pecuniarie "ordinarie" dell'articolo 23, in forma aggravata, con un aumento sanzionatorio di un terzo rispetto all'importo *«previsto per la violazione originaria»* (ma sarebbe stato più adeguato far riferimento all'importo "applicabile" per la violazione originaria, dal momento che – com'è noto – le circostanze aggravanti non operano sulle cornici edittali, ma incidono sulla sanzione concreta), comunque sempre nell'ambito dei limiti massimi edittali definiti dall'articolo 23.

Si noti, per ultimo, che, secondo quanto specificato dall'articolo 22, comma 3 d.lgs. 179/2026, il riconoscimento della potestà di applicare misure non pecuniarie in materia di intelligenza artificiale non interferisce con i poteri sanzionatori "generalisti" riconosciuti alle varie autorità (richiamate dall'articolo 20, comma 2 l. 132/2025, ossia, nell'ordine, CONSOB, IVASS, AgID, ACN) *«in base alla legislazione di settore»*.

Le sanzioni amministrative «*pecuniarie*» hanno un campo di applicazione specificamente disegnato, nel rispetto del generale principio di tassatività posto dall'articolo 1 l. 689/1981, dalle disposizioni del d.lgs. 179/2026 (articolo 23), secondo un'articolazione direttamente riflessa dai criteri dell'articolo 99, parr. 3 ss. Al Act.

Si osserva preliminarmente, tuttavia, che il campo di applicazione delle sanzioni amministrative pecuniarie, in concreto, è potenzialmente soggetto a un duplice fenomeno di erosione:

- nei casi di applicabilità delle ipotesi di “scarsa offensività” che siano ritenute sanzionabili con misure non pecuniarie ai sensi dell'articolo 22 d.lgs. 179/2026, le quali sono suscettibili di ritagliare così un autonomo spazio per la fattispecie dell'illecito amministrativo “tenue” in materia di intelligenza artificiale;
- nei casi in cui una fattispecie di reato sia destinata a prevalere, in forza della clausola di sussidiarietà espressa inserita nei vari commi dell'articolo 23 («*Salvo che il fatto costituisca reato*»), poiché punisce una forma di offesa ulteriore o più grave rispetto al mero illecito amministrativo previsto per la violazione dell'Al Act, di cui esaurisce integralmente il disvalore.

Al di là delle specifiche cornici edituali previste per i vari illeciti amministrativi, ricavate (nel massimo) nel rispetto dei limiti stabiliti dall'articolo 99 Al Act, sia consentito osservare come il d.lgs. 179/2026 propone un “adeguamento” letterale alla disposizione sanzionatoria dell'articolo 99 Al Act.

Infatti, mentre nel regolamento europeo l'alternatività tra criterio numerico (milioni di euro) e criterio percentuale (fatturato mondiale totale annuo) – come detto – trova un'espressa discriminante nella natura di «*impresa*» del soggetto responsabile della violazione e nella maggiore afflittività in concreto della sanzione calcolata su base percentuale («*o, se l'autore del reato è un'impresa, fino al [...] del fatturato mondiale totale annuo dell'esercizio precedente, se superiore*»), il legislatore nazionale tiene fermo il metodo di calcolo della sanzione basato sul criterio del “maggior importo” omettendo di richiamare esplicitamente, quale soggetto sanzionato, l'«*impresa*».

La mancanza, almeno a una prima lettura, non sembra avere un impatto in termini di conseguenze sanzionatorie per i soggetti trasgressori e obbligati in solido al pagamento delle sanzioni pecuniarie, essendo la mancata replica della lettera dell'articolo 99 Al Act compensata dal riferimento al «*fatturato mondiale totale annuo dell'esercizio precedente*», che si può ritenere implicitamente richiami la condizione applicativa della natura di «*impresa*» del soggetto sanzionato.

4. Intelligenza artificiale e diritto penale. Il delitto di «*Omessa adozione di misure di sicurezza nei sistemi di intelligenza artificiale e alterazione illecita dei sistemi*» (articolo 437-bis c.p.).

Premesso tale quadro di sintesi della normativa eurounitaria di riferimento e muovendo verso la direttrice di attuazione (*rectius*, adeguamento) interna seguita dal Governo, occorre volgere lo sguardo – per i profili penali sostanziali – alle disposizioni del d.lgs. 160/2026.

Il Titolo II del provvedimento, in particolare, reca «*Disposizioni penali, sostanziali e processuali, in materia di realizzazione e impiego illeciti dei sistemi di intelligenza artificiale e disposizioni processuali civili in materia di risarcimento dei danni cagionati dall'utilizzo dei medesimi sistemi*», nell'ambito del quale figura un Capo I dedicato alle «*Disposizioni sanzionatorie e in materia di procedimento penale*».

In tale contesto – in cui trovano applicazione, per espressa previsione normativa (articolo 11), le definizioni già contenute nell'articolo 3 AI Act, che dunque assumeranno diretto rilievo quali norme extrapenali integratrici del precetto penale – l'articolo 12 d.lgs. 160/2026, nell'esercizio della richiamata delega di cui all'articolo 24, comma 5, lett. b) l. 132/2025, introduce, all'interno del Titolo VI del Codice penale (dedicato ai delitti contro l'incolumità pubblica), nel capo I relativo ai «*delitti di comune pericolo mediante violenza*», un nuovo delitto di «*Omessa adozione di misure di sicurezza nei sistemi di intelligenza artificiale e alterazione illecita dei sistemi*» (articolo 437-bis c.p.).

Da un punto di vista generale, anzitutto, occorre segnalare la scelta del legislatore delegato di optare per una politica di incriminazione *soft* degli illeciti connessi ai sistemi di intelligenza artificiale, riservando (almeno per il momento) allo stigma della sanzione penale soltanto specifiche e selezionate categorie di condotte, per converso mantenendo nel perimetro degli illeciti amministrativi e delle relative sanzioni (non pecuniarie e pecuniarie) tutte le altre multiformi violazioni della disciplina regolatoria di settore.

Simile opzione, che segue la direzione di un diritto penale "minimo" – al netto della *vexata quaestio*, suscettibile di riproporsi anche rispetto alla disciplina sanzionatoria in tema di intelligenza artificiale, circa la qualificazione degli illeciti amministrativi di settore come illeciti "punitivi", di natura sostanzialmente penale (riconducibili al campo della *matière pénale* e alle garanzie eurounitarie e convenzionali, tra le quali il divieto di *bis in idem* degli articoli 50 CDFUE e 4 Protocollo 7 CEDU, per cui *infra*) – appare oltremodo significativa, trovando una condivisibile *ratio* giustificativa nelle note e condivise difficoltà, ben messe in luce dall'ormai copiosa letteratura specialistica, che si frappongono alla costruzione di un autonomo statuto della responsabilità penale "da intelligenza artificiale".

Altrettanto peculiare – ma parimenti comprensibile, alla luce dei beni e degli interessi potenzialmente incisi dalla fattispecie di reato – risulta l'introduzione

della nuova disposizione incriminatrice nell'ambito dei delitti contro l'incolumità pubblica (Titolo VI del Codice penale), ossia in un "contesto" di tutela riferito a fatti caratterizzati dalla diffusività del danno o del pericolo verso un numero non determinabile *a priori* di persone – che, almeno in prima battuta, suona, se non estraneo, alquanto "lontano" dalla comune percezione circa l'impatto, in termini di oggettività giuridiche, dei sistemi di intelligenza artificiale.

Tra l'altro – come meglio si vedrà nella successiva analisi delle fattispecie – i reati previsti dall'articolo 437-*bis* c.p. assumono natura plurioffensiva, in considerazione del fatto che, tra i beni presidiati dall'incriminazione, figurano anche la vita e l'incolumità individuale.

Sotto questo profilo – come spiegato nella Relazione illustrativa dello schema di decreto legislativo – *«[l]a collocazione tra i delitti di comune pericolo mediante violenza (artt. 422-437 c.p.) riflette il bene giuridico tutelato e l'affinità strutturale con l'art. 437, che disciplina l'omessa collocazione di impianti destinati a prevenire infortuni sul lavoro»*.

Sarebbe parsa più congeniale, forse – nell'ambito di un tentativo maggiormente ambizioso di codificazione di una "parte speciale" dei reati da intelligenza artificiale – l'introduzione della disposizione in un autonomo titolo del Codice penale, dedicato alla responsabilità penale connessa ai sistemi di intelligenza artificiale, tenuto altresì conto dei criteri direttivi della l. 132/2025, che richiedeva addirittura la specificazione di *«criteri di imputazione della responsabilità penale delle persone fisiche [...] per gli illeciti inerenti a sistemi di intelligenza artificiale, che tenga conto del livello effettivo di controllo dei sistemi predetti da parte dell'agente»* (articolo 24, comma 5, lett. c)), con riferimento dunque a una "parte generale" della disciplina penale (ad esempio, nell'ambito di un ipotetico comma 1-*bis* dell'articolo 42 c.p., in coda alla disposizione sulla *suitas* della condotta, o di un articolo 42-*bis* c.p.). La disposizione risultante dall'esercizio della delega legislativa, pertanto, risente direttamente – per struttura e per prospettive applicative – dei fattori limitativi poc'anzi evidenziati.

Il reato, come si vedrà meglio tra poco, è articolato in un duplice "nucleo" di tipicità penale – condensato nei commi 1, 3 e 4, da un lato, e nel comma 2, dall'altro – che assoggetta a incriminazione condotte tra di loro eterogenee, per disvalore e per materialità, che per tale ragione meritano un'analisi separata.

4.1. L'omessa adozione di misure di sicurezza e di sorveglianza umana per i sistemi di intelligenza artificiale ad alto rischio (articolo 437-*bis*, commi 1 e 3 c.p.).

L'ipotesi del comma 1 costituisce reato di pericolo, preordinato a colpire la "filiera" (o, com'è definita anche dall'AI Act, la *«catena del valore»*) dell'intelligenza artificiale nelle sue varie articolazioni, comprensive delle fasi

di «*progettazione*», di «*addestramento*», di «*produzione*» e di «*immissione sul mercato*».

Si nota, preliminarmente, che la disposizione contenuta dell'articolo 437-*bis*, comma 1 c.p., come introdotta dal d.lgs. 160/2026 – a differenza della versione contenuta nello schema di decreto legislativo – non contempla più la fase dell'«*utilizzo professionale*», la quale, come oltre si noterà, è stata nondimeno autonomamente considerata nel “nuovo” (rispetto allo schema) comma 4 dell'articolo.

Dal punto di vista del soggetto agente, dunque, la fattispecie criminosa del comma 1 è solo apparentemente configurata come reato comune («*Chiunque*»), ma ha, in realtà, natura propria, dal momento che le condotte incriminate sono correlate alla violazione di obblighi il cui adempimento è posto a carico di specifici soggetti “qualificati”, secondo le disposizioni della normativa regolatoria in materia di intelligenza artificiale poco sopra illustrata.

In proposito, come osservato nella Relazione illustrativa allo schema di decreto legislativo, «*[l]a condotta è imputata a chi opera in una delle fasi del ciclo di vita rilevanti*

(progettazione, addestramento, produzione, immissione sul mercato o utilizzo professionale)», attraverso un elenco «*funzionale a coprire tutte le posizioni di garanzia individuate dal regolamento (UE) 2024/1689 – fornitore, importatore, distributore, deployer*».

L'identificazione dei soggetti attivi del reato, tuttavia, richiede una più attenta ponderazione.

Se, da un lato, l'attività di «*progettazione*» e di «*produzione*» consentono di ricomprendere certamente tra i soggetti attivi del reato il *provider* (il quale, secondo la definizione dell'articolo 3, par. 1, n. 3 Al Act, «*sviluppa*» o «*fa sviluppare*» un sistema di intelligenza artificiale), dall'altro, l'attività di «*immissione sul mercato*» rende punibile per il reato lo stesso *provider* e l'importatore (soggetti ai quali fa capo la «*immissione sul mercato*» del sistema, intesa quale «*prima messa a disposizione*» del sistema sul mercato dell'Unione, ai sensi dell'articolo 3, par. 1, n. 9 Al Act).

Un discorso a parte merita la posizione del *deployer*, la quale, come detto, espressamente considerata nella versione del comma 1 prevista nello schema di decreto legislativo (in relazione alla fase di «*utilizzo*»), oggi è stata (almeno apparentemente, *infra*) espunta dal campo applicativo della fattispecie penale in esame, ma è stata contestualmente recuperata – con le limitazioni che oltre si diranno – attraverso l'introduzione di un'autonoma fattispecie dedicata nel successivo comma 4, che vede quale proprio soggetto attivo l'«*utilizzatore professionale*» (ossia proprio il soggetto che, alla luce dell'articolo 3, par. 1, n. 4 Al Act, «*utilizza*» il sistema al di fuori delle ipotesi di «*un'attività personale non professionale*»).

Tra i vari soggetti della filiera – diversamente da quanto riferito dal legislatore delegato nella Relazione illustrativa – non parrebbe direttamente impattato da profili di responsabilità penale, in virtù dell’espresso rinvio alle definizioni di cui all’articolo 3 AI Act, il distributore, il quale, a rigore, «*mette a disposizione*» il sistema (e dunque, realizza «*la fornitura [...] per la distribuzione o l’uso sul mercato dell’Unione nel corso di un’attività commerciale*», come previsto dall’articolo 3, par. 1, n. 10 AI Act) e, pertanto, non realizza la «*immissione sul mercato*» del sistema – la sola condotta sanzionata dall’articolo 437-bis c.p. – intesa quale «*prima messa a disposizione*».

Il basilare canone di tassatività del precetto penale, dunque, parrebbe impedire un’interpretazione analogica *in malam partem* della condotta incriminata, in senso ostativo alla ricomprensione del distributore nel novero dei soggetti attivi del reato, a pena di un’interpretazione a concreto rischio di censura di costituzionalità.

Più sfumata è la posizione del «*rappresentante autorizzato*», il quale, in quanto soggetto incaricato dal *provider* di «*adempiere ed eseguire per suo conto gli obblighi e le procedure*» previste dall’AI Act, astrattamente potrebbe concorrere nel reato (articolo 110 c.p.) con il *provider*, in ragione della condivisione degli obblighi (e della relativa posizione di garanzia) di derivazione contrattuale.

Inoltre, ai fini della completezza dell’analisi, devono essere considerati anche i margini di potenziale responsabilità del «*fornitore a valle*» (c.d. *downstream provider*), che, quale «*fornitore di un sistema di IA [...] che integra un modello di IA*» (dell’articolo 3, par. 1, n. 68 AI Act), potrebbe considerarsi coinvolto nelle fasi di «*progettazione*» e di «*produzione*», intese in senso lato, del sistema.

Meno agevole, invece, è la ricostruzione del perimetro di rilevanza penale dell’attività di «*addestramento*», la quale – è importante evidenziarlo – riguarda, a rigore, i modelli di intelligenza artificiale e, solo per via derivata, i relativi sistemi ad alto rischio che quel modello incorporano.

In via preliminare, si osserva che la nozione di «*addestramento*», a un’attenta analisi del dato normativo, non trova un’espressa codificazione, né uno specifico statuto regolatorio all’interno dell’AI Act, il quale si limita a richiamarla, oltre che in vari considerando, all’interno di specifiche disposizioni settoriali, senza tuttavia individuare in modo univoco uno o più soggetti “gestori” di tale fase del ciclo di vita del modello (e del sistema che utilizza quel modello).

Tanto premesso, nella prassi operativa, il concetto di «*addestramento*» è generalmente inteso in senso ampio, comprensivo non soltanto dell’addestramento in senso stretto (*training*, cioè l’addestramento iniziale e le attività a esso collaterali e accessorie, quali selezione dei *dataset*, raccolta dei dati di *training*, pulizia e normalizzazione dei dati, etichettatura dei dati),

ma anche delle attività di *pre-training* (intesa quale fase iniziale soprattutto del *machine learning*, in cui il modello apprende informazioni e schemi su base massiva, prima di essere oggetto di *tuning* per specifici compiti), di *fine tuning* (inteso come ulteriore addestramento di un modello già esistente per adattarlo alla finalità specifica impressa dal *provider* o dal *deployer*) e di *retraining* (ossia di aggiornamento del modello attraverso nuovi o più recenti dati, in modo da mantenere o migliorarne la *performance*).

Dall'ampio perimetro concettuale descritto – nella prospettiva di identificazione del soggetto attivo del reato di cui all'articolo 437-*bis* comma 1 c.p. – è possibile dedurre i soggetti che possono risultare coinvolti nella fase di addestramento, i quali sono in prima battuta certamente riconducibili, almeno, al *provider* (e al *downstream provider*) e al *deployer*.

Rispetto a quest'ultimo, la sua riconducibilità anche all'attività di «addestramento» e la conseguente imputabilità a suo carico di omissioni relative a tale fase, punibili ai sensi del comma 1, crea non poche incognite applicative – come oltre si vedrà – in considerazione dell'autonoma previsione di una responsabilità dell'«*utilizzatore professionale*» nel comma 4, in relazione a una fattispecie la quale, tuttavia, risulta diversamente connotata dal punto di vista oggettivo (fatto tipico) e soggettivo (elemento psicologico).

Tra le disposizioni che regolano l'attività di addestramento si menziona, per particolare rilevanza ai fini della condotta incriminata, l'articolo 10 AI Act, che prevede i criteri di qualità dei *set* di dati di addestramento dei sistemi ad alto rischio, nonché pratiche di *governance* e di gestione dei dati.

Un punto assai rilevante e sensibile, infine, attiene alla possibile diffusione o "migrazione" delle responsabilità penali (ma pari discorso potrebbe valere per le responsabilità di matrice amministrativa) al di fuori e al di là della filiera dell'intelligenza artificiale in senso stretto, verso soggetti "terzi" rispetto ai soggetti "qualificati" previsti dall'AI Act – per il tramite della clausola generale in materia di concorso di persone nel reato (articolo 110 c.p.) – che agiscano in qualità di *extraneus* nella commissione della fattispecie propria.

Si pone, in questo senso, ad esempio, il tema – che, a ben vedere, costituisce un *leitmotiv* della giurisprudenza e della letteratura in materia di diritto penale d'impresa – della rilevanza penale dell'attività di eventuali consulenti esterni che supportino, a livello tecnico o intellettuale, una società che agisce nelle varie "fasi di vita" del sistema di intelligenza artificiale ad alto rischio come *provider* o *deployer* (quest'ultimo, in relazione al solo addestramento), fornendo un contributo – quanto meno agevolatore nella prospettiva dell'articolo 110 c.p. – rispetto alla definizione, all'implementazione o all'adeguamento delle misure tecniche "di sicurezza" di un sistema di intelligenza artificiale ad alto rischio richiamate dall'articolo 437-*bis*, comma 1 c.p.

Beninteso, l'ipotetica responsabilità concorsuale dell'*extraneus*, almeno per quanto riguarda la fattispecie dolosa del comma 1, imporrà il riscontro e l'accertamento di uno specifico quoziente psicologico doloso, secondo la duplice accezione del "dolo di concorso", inteso quale rappresentazione e volontà del fatto e di consapevolezza del proprio contributo causale rispetto alla condotta criminosa altrui.

Stante il contesto operativo in cui normalmente si colloca il contributo di natura tecnica o intellettuale del "terzo" – che ragionevolmente interverrà nell'ambito di un'attività imprenditoriale "lecita di base" – è più verosimile ipotizzare spazi di applicabilità per un contributo concorsuale dell'*extraneus* rispetto alla fattispecie (come si vedrà) colposa del comma 3, rispetto alla quale può astrattamente trovare attivazione la disciplina in tema di cooperazione colposa (articolo 113 c.p.), la quale presuppone un contegno soggettivo meno pregnante, articolato, da un lato, sulla coscienza e volontà di concorrere all'altrui condotta violatrice delle norme cautelari e, dall'altro, la prevedibilità ed evitabilità dell'evento lesivo.

La tipicità dell'incriminazione in esame, peraltro, trova un significativo limite applicativo nel proprio oggetto materiale, che abbraccia non qualsiasi sistema di intelligenza artificiale (in relazione ai vari livelli di rischio in precedenza esaminati), ma soltanto i «*sistemi di intelligenza artificiale ad alto rischio*».

La scelta del legislatore delegato, se apparentemente sembra porsi in antinomia con i principi della legge delega – che all'articolo 24, comma 5, lett. b) fa generico e onnicomprensivo riferimento ai «*sistemi di intelligenza artificiale*», non limitando il cono della tutela penalistica ai soli sistemi di intelligenza artificiale ad alto rischio – a un esame più approfondito risulta coerente con i dettami della l. 132/2025.

Infatti, la richiamata disposizione dell'articolo 24, comma 5, lett. b) l. 132/2025, esigendo l'incriminazione di fattispecie di reato che siano «*incentrate sull'omessa adozione o sull'omesso adeguamento di misure di sicurezza*», implicitamente sottende che oggetto della tutela penale siano i soli sistemi di intelligenza artificiale ad alto rischio, poiché gli unici per i quali sono previsti – all'interno degli articoli 8 ss. Al Act – specifici requisiti aventi la finalità di garantire la "sicurezza" del sistema (ad esempio, per la qualità, l'accuratezza e la rappresentatività dei dati di addestramento, la supervisione umana, l'accuratezza, la robustezza e la cibersicurezza durante tutto il ciclo di vita) e, per tale via, dei beni giuridici identificati dalla fattispecie incriminatrice.

Si noti che sono formalmente esclusi dalla tutela penalistica i modelli di intelligenza artificiale per finalità generali (GPAI), per i quali – come visto in precedenza – in relazione al loro livello di rischio (basico o sistemico), sarebbero previsti dei requisiti specifici riconducibili al novero delle «*misure*

di sicurezza», ma, in quanto “modelli” e non “sistemi”, esulano dal fuoco della fattispecie incriminatrice.

Nondimeno, tali modelli, nel momento e nella misura in cui sono inglobati all’interno di un sistema di intelligenza artificiale ad alto rischio, refluiscano dentro lo spettro dell’incriminazione, con riferimento alle ipotesi in cui i rispettivi presidi di “sicurezza” non siano correttamente adottati, implementati o adeguati.

Si noti, in proposito, che proprio la perimetrazione dell’oggetto materiale della tutela penale ai soli sistemi di intelligenza artificiale ad alto rischio costituisce un punto di centrale rilevanza sotto il profilo dell’immediata applicabilità delle fattispecie in esame, da parte degli uffici giudiziari, all’indomani della data di entrata in vigore (30 settembre 2026) del d.lgs. 160/2026.

In particolare, per le sole ipotesi di reato dei commi 1, 3 e 4 – che chiamano in causa violazioni configurabili in termini di omesso adempimento (totale o parziale) di obblighi normativi posti, in prima battuta (ma non solo, come si vedrà *infra*), dalla normativa regolatoria dell’AI Act – occorre tenere in considerazione il regime temporale di cogenza dei requisiti previsti dall’AI Act per i sistemi di intelligenza artificiale ad alto rischio, stabilito e di recente modificato dal legislatore europeo (*retro*, paragrafo 2).

Il rinvio al 2 dicembre 2027 (per quanto riguarda i sistemi di cui all’Allegato III) e al 2 agosto 2028 (per quanto riguarda i sistemi di cui all’Allegato I) per l’applicabilità delle disposizioni dell’AI Act relative ai sistemi di intelligenza artificiale ad alto rischio (classificazione, requisiti, obblighi dei fornitori, dei *deployer* e di altre parti), infatti, “trascina” con sé anche l’impossibilità per il pubblico ministero e per il giudice penale, per i fatti commessi successivamente alla data di entrata in vigore del d.lgs. 160/2026 (30 settembre 2026) ma prima delle suddette scadenze temporali previste dall’AI Act, rispettivamente, di iscrivere la notizia di reato nel registro dell’articolo 335 c.p.p., di avanzare un’imputazione, nonché di pronunciare sentenza di condanna per le fattispecie sopra indicate.

In altri termini, a una prima analisi, parrebbe potersi affermare che le suddette incriminazioni siano destinate a operare “a scoppio ritardato”.

Diverso discorso deve essere fatto per l’ipotesi dell’articolo 437-*bis*, comma 2 c.p., la quale – come oltre si analizzerà – non chiama in causa l’osservanza dei requisiti cogenti previsti per i sistemi di intelligenza artificiale ad alto rischio, i quali, in tale contesto, rappresentano il mero *target* dell’azione criminosa, siano o meno in linea con i requisiti regolatori in tema di misure tecniche di sicurezza e di sorveglianza umana.

Tuttavia, la provvisoria conclusione circa l’effetto di “congelamento” delle fattispecie delittuose punite dall’articolo 437-*bis*, commi 1, 3 e 4, fino alle date stabilite dal legislatore europeo in cui assumeranno piena efficacia per gli operatori gli obblighi in materia di sistemi di intelligenza artificiale ad alto

rischio, deve essere rivista – quanto meno per le ipotesi dei commi 1 e 3 – all’esito della più attenta analisi delle condotte ivi incriminate, sotto il profilo dell’individuazione degli obblighi cogenti (ulteriori rispetto a quelli previsti dall’AI Act) che devono ritenersi applicabili per i sistemi di intelligenza artificiale ad alto rischio.

In quest’ultima prospettiva, che sarà esaminata nel seguito, le incriminazioni neointrodotte dal d.lgs. 160/2026 costituirebbero un armamentario sanzionatorio già “pronto all’uso” nei confronti degli operatori della filiera dell’intelligenza artificiale.

Poste queste coordinate di carattere generale, è possibile ora procedere all’analisi delle diverse fattispecie di reato contemplate dall’articolo 437-*bis* c.p.

La condotta tipica del comma 1 è costruita alla stregua di un reato omissivo proprio ed è imperniata – secondo il paradigma della norma a più fattispecie (o norma mista alternativa) – sull’omissione di «*misure tecniche di sicurezza [...] idonee a prevenire malfunzionamenti o alterazioni del funzionamento dei sistemi*» ovvero sull’omissione di «*misure di sorveglianza umana*».

La disposizione, a una prima lettura, si caratterizza per l’estrema ampiezza della terminologia usata per tipizzare le condotte – che apre possibili spazi di censura sotto il profilo della mancanza di precisione/determinatezza del fatto, in relazione al complesso definitorio dell’AI Act (e delle altre normative di settore eventualmente rilevanti nel caso concreto, *infra*) – e, per tale ragione, impone necessarie precisazioni finalizzate a chiarire la reale portata della norma incriminatrice.

Anzitutto, si rileva che, mentre la l. 132/2025 fa riferimento a condotte omissive correlate a non meglio qualificate «*misure di sicurezza*», il d.lgs. 160/2026 – come visto – dettaglia autonomamente, senza che ciò fosse previsto o imposto dalla legge delega, il novero di tali misure, limitando la tutela penale, in chiave di specialità per specificazione rispetto al dettato della l. 132/2025, alle sole «*misure tecniche di sicurezza [...] idonee a prevenire malfunzionamenti o alterazioni del funzionamento dei sistemi*».

Ciò premesso, un punto di assoluta centralità nell’analisi della fattispecie – anche nella duplice chiave, preventiva (di identificazione e di valutazione dei rischi) e “frappositiva” (di definizione dei protocolli di mitigazione del rischio reato), del modello di organizzazione, gestione e controllo previsto dal d.lgs. 231/2001 (*infra*, paragrafo 5) – attiene all’esatta identificazione dell’oggetto dell’omissione attribuibile al soggetto agente, ossia le «*misure tecniche di sicurezza previste per la progettazione, l’addestramento, la produzione, l’immissione sul mercato di sistemi di intelligenza artificiale ad alto rischio, idonee a prevenire malfunzionamenti o alterazioni del funzionamento dei sistemi*».

L'ampia terminologia utilizzata dalla disposizione dell'articolo 437-bis, comma 1 c.p. (a cui rinvia anche il comma 3), infatti, non rende immediatamente comprensibile:

- a) quali presidi debbano essere considerati ai fini penalistici, per i sistemi di intelligenza artificiale ad alto rischio, come «*misure tecniche di sicurezza [...] idonee a prevenire malfunzionamenti o alterazioni del funzionamento dei sistemi*»;
- b) ancor prima, quali siano i plessi normativi di riferimento nell'ambito dei quali devono essere «*previste*» – e reperite dall'interprete, ivi compreso il giudice penale – le suddette misure tecniche di "sicurezza", in termini di obblighi positivi di condotta posti in capo ai soggetti della filiera rispetto ai sistemi di intelligenza artificiale ad alto rischio, il cui inadempimento informa il rimprovero doloso (o colposo) a carico dell'agente;

Per fornire una risposta ai due quesiti, posti in rapporto di stretta interdipendenza logica, occorre preliminarmente osservare che il testo dell'articolo 437-bis, comma 1 c.p., a ben vedere, oltre a non meglio specificare il novero delle «*misure tecniche di sicurezza*» recanti i requisiti ivi indicati, appare del tutto "neutro" anche sotto il profilo dell'individuazione delle normative extrapenalistiche di riferimento impositive degli obblighi tecnici di "sicurezza" e integratrici del precetto penale, limitandosi a connotare tali obblighi in relazione al loro "contesto" (ossia deve trattarsi di obblighi da adempiere per le fasi di progettazione, di addestramento, di produzione e di immissione sul mercato) e alla loro finalità/funzionalità (ossia deve trattarsi di obblighi che garantiscano di «*prevenire malfunzionamenti o alterazioni del funzionamento dei sistemi*»).

Il primo e più immediato referente normativo per la determinazione degli obblighi in tema di «*misure tecniche di sicurezza [...] per la progettazione, l'addestramento, la produzione, l'immissione sul mercato di sistemi di intelligenza artificiale ad alto rischio*» – abbastanza intuitivamente – non può che rinvenirsi nella disciplina "quadro" in materia di intelligenza artificiale rappresentata dall'AI Act, il quale, come sommariamente delineato nei precedenti paragrafi, pone un denso articolato di disposizioni direttamente ed esplicitamente regolatorie, per le loro varie fasi di vita, dei sistemi di intelligenza artificiale (anche ad alto rischio).

Nondimeno, già rispetto alla disciplina dell'AI Act, il *gap* descrittivo dell'articolo 437-bis, comma 1 c.p. appare piuttosto consistente, non essendo questione di pronta e agevole soluzione individuare quali tra i numerosi requisiti previsti dall'AI Act per i sistemi di intelligenza artificiale ad alto rischio siano destinati a fungere, nella prospettiva penalistica, da «*misure tecniche di sicurezza [...] idonee a prevenire malfunzionamenti o alterazioni del funzionamento dei sistemi*», dal momento che, tra l'altro, non sussiste una corrispondenza biunivoca tra la terminologia utilizzata ai fini dell'incriminazione dall'articolo 437-bis, comma 1 c.p. e le definizioni o,

comunque, le altre disposizioni di carattere regolatorio puro contenute nell'AI Act.

In tal senso, può icasticamente dirsi che l'articolo 437-*bis*, comma 1 c.p. e l'AI Act "non parlano la stessa lingua".

Tuttavia, essendo compito dell'interprete attribuire alla norma un significato che ne consenta l'applicazione, occorre guardare al tessuto regolatorio dell'AI Act per individuare quali, tra i vari requisiti e presidi imposti agli operatori della filiera per i sistemi di intelligenza artificiale ad alto rischio, siano sussumibili nel novero delle suddette «*misure tecniche di sicurezza*», come ulteriormente connotate dal legislatore delegato.

Ciò presuppone, però, un ulteriore preliminare chiarimento in merito alla finalità di tali misure – che, secondo espressa previsione del comma 1, devono essere funzionali «*a prevenire malfunzionamenti o alterazioni del funzionamento dei sistemi*» – non essendovi nemmeno in tal caso, né nell'AI Act, né nella l. 132/2025, né nel d.lgs. 160/2026, una definizione espressa delle nozioni di "malfunzionamento" e di "alterazione del funzionamento".

In assenza di precisi referenti normativi e tenendo conto delle dinamiche tecnico-operative dei sistemi di intelligenza artificiale, è presumibile che le nozioni di "malfunzionamento" e di "alterazione del funzionamento" saranno interpretate – con il surrettizio obiettivo di garantire più ampie *chance* di successo alla fattispecie incriminatrice – in senso estensivo, atto a ricomprendere tutte le disfunzioni di natura progettuale e applicativa del sistema, che trovino la propria matrice in cause endogene (condotta dello stesso soggetto "qualificato") e – soprattutto per quanto concerne la "alterazione del funzionamento" – esogene (condotta di un soggetto terzo). Quanto alle cause endogene, il dubbio che sorge, in sede di prima analisi, è se tra le fattispecie di "malfunzionamento" – riconducibili a una matrice interna all'organizzazione che sviluppa (e non più, data la differenziazione dell'area di rilevanza penale oggi introdotta nel comma 4, che utilizza) il sistema di intelligenza artificiale ad alto rischio – rientrino i fenomeni tipici di "allucinazione" (che ricorre nel caso in cui il sistema produca un *output*, presentato come vero e affidabile, che in realtà è privo di fondamento nei fatti o nelle fonti sottostanti) e di *bias* (intesa come distorsione sistematica negli *output*, che riflette squilibri o pregiudizi presenti nei dati di addestramento o nel modo in cui il modello è stato costruito) dei modelli integrati nei sistemi di intelligenza artificiale.

Ove la risposta fosse positiva, l'esenzione da responsabilità penale per la fattispecie – con particolare riferimento alla sua versione colposa del comma 3 – esigerebbe, ad esempio, dal *provider* lo studio, l'identificazione e la successiva adozione, in fase di sviluppo (nello specifico, di addestramento) del modello sottostante al sistema di intelligenza artificiale ad alto rischio, di misure e di presidi volti a garantire che i *set* di dati possiedano «*le proprietà statistiche appropriate, anche per quanto riguarda le persone o i gruppi di*

persone in relazione ai quali il sistema di IA ad alto rischio è destinato a essere usato, prestando particolare attenzione all'attenuazione di possibili distorsioni nei set di dati, suscettibili di incidere sulla salute e sulla sicurezza delle persone, di avere un impatto negativo sui diritti fondamentali o di comportare discriminazioni vietate dal diritto dell'Unione» (considerando n. 67).

Si noti, tra l'altro, che la condotta omissiva in esame è ulteriormente specificata dal legislatore delegato attraverso il riferimento al requisito di "idoneità" («*idonee*») delle «*misure tecniche di sicurezza*».

Non è chiaro, tuttavia, quale sia il ruolo e il significato di tale requisito di fattispecie e, in particolare, se esso debba essere inteso come un ulteriore elemento costitutivo della condotta incriminata, con potenziale portata estensiva della sua applicazione.

Una prima ipotesi ricostruttiva consente di interpretare in senso neutro l'aggettivo «*idonee*», in quanto riferito, in rapporto di mezzo a fine, all'obiettivo preventivo della misura tecnica di sicurezza (dunque, quale sinonimo di «*finalizzate a*»).

Una seconda lettura, invece, potrebbe annettere a tale aggettivo un significato sostanziale, riconoscendovi una funzione realmente costitutiva della fattispecie penale, tale per cui il reato potrebbe ricorrere non soltanto quando le misure tecniche di sicurezza siano state omesse *tout court* (omissione totale), ma anche nel caso in cui delle misure tecniche di sicurezza siano state implementate da parte dell'organizzazione, ma non siano «*idonee*» (in primo luogo, dal punto di vista tecnico) rispetto alla finalità preventiva di malfunzionamenti o di alterazioni.

Seguendo questa opzione, peraltro, si osserva che il requisito di "idoneità" – inteso alla stregua di "adeguatezza" – non è ulteriormente connotato dal legislatore delegato, dal momento che non è individuato uno specifico parametro da spendere per il giudizio di relazione che esso implica, in assenza, allo stato attuale, di una normazione di secondo livello (ad esempio, di "linee guida" sui requisiti dei sistemi di intelligenza artificiale ad alto rischio) emessa da parte delle competenti autorità europee (tra cui il c.d. *AI Office*).

Ad ogni modo, occorre osservare, altresì, che il requisito di idoneità-adeguatezza non sarebbe scevro da criticità applicative, giacché, ove ritenuto sussistente, sarebbe necessario sciogliere il nodo interpretativo circa la sua caratterizzazione in astratto ovvero in concreto.

In questo senso, cioè, potrebbe in futuro porsi il quesito se un soggetto della catena del valore dell'intelligenza artificiale – rientrando nel novero dei soggetti attivi dell'articolo 437-bis, comma 1 c.p. – possa andare esente da responsabilità penale semplicemente adottando le "misure tecniche di sicurezza" corrispondenti agli *standard* normativi o indicati dalle *best practice* di settore ovvero, diversamente, debba condurre ulteriori valutazioni e

verifiche, caso per caso, sull'efficacia e sull'effettività di una determinata misura rispetto al rischio di malfunzionamento o di alterazione del sistema. Certo è che, qualora si adotti quest'ultima lettura, si aprirebbero ampi e indiscriminati spazi per il riconoscimento, *ex post*, nella fase "patologica" dell'accertamento giudiziale, della responsabilità penale degli operatori di settore, secondo la nota – seppur vituperata – logica del *post hoc propter hoc*, ben conosciuta nel settore degli illeciti colposi come distorsione retrospettiva del "senno di poi" (*hindsight bias*), mediante la quale si tende a ritenere prevedibile un evento, una volta che l'evento è ormai noto, e a costruire di conseguenza la regola cautelare che avrebbe potuto impedirlo.

Al netto di queste ulteriori problematiche interpretative, riprendendo le fila del discorso inerente alla definizione dello spettro applicativo della fattispecie penale in relazione agli obblighi previsti dall'AI Act, per i sistemi di intelligenza artificiale ad alto rischio, potrebbero essere ricomprese nel novero delle «*misure tecniche di sicurezza*» nei termini indicati dall'articolo 437-bis, comma 1 c.p., tra le altre:

- l'adozione di un «*sistema di gestione dei rischi*» (articolo 9 AI Act), comprensivo in particolare della «*identificazione e analisi dei rischi noti e ragionevolmente prevedibili per la salute, la sicurezza e i diritti fondamentali*» (par. 2, lett. a), nonché «*misure di gestione dei rischi opportune e mirate*» volte a mitigare i rischi predetti (par. 2, lett. d);
- in fase di addestramento, il rispetto di «*criteri di qualità*» per i «*set di dati di addestramento, convalida e prova*» (articolo 10 AI Act), tra i quali rientrano le «*pratiche di governance e gestione dei dati adeguate alla finalità prevista*» del sistema (parr. 1 e 2), nonché «*un esame atto a valutare le possibili distorsioni suscettibili di incidere sulla salute e sulla sicurezza delle persone, di avere un impatto negativo sui diritti fondamentali o di comportare discriminazioni vietate dal diritto dell'Unione*» (par. 2, lett. f) e «*misure adeguate per individuare, prevenire e attenuare le possibili distorsioni*» appena descritte (par. 2, lett. g);
- l'adozione di misure preordinate a «*conseguire un adeguato livello di accuratezza, robustezza e cibersecurity*» (articolo 15, par. 1 AI Act) e, in particolare, di «*misure tecniche e organizzative*» volte a rendere i sistemi di intelligenza artificiale ad alto rischio «*il più resilienti possibile per quanto riguarda errori, guasti o incongruenze che possono verificarsi all'interno del sistema o nell'ambiente in cui esso opera, in particolare a causa della loro interazione con persone fisiche o altri sistemi*» ovvero, ancora, di «*soluzioni tecniche di ridondanza, che possono includere piani di backup o fail-safe*» (articolo 15, par. 4 AI Act);
- in relazione ai sistemi di intelligenza artificiale ad alto rischio «*che proseguono il loro apprendimento dopo essere stati immessi sul*

mercato o messi in servizio», l'adozione di misure tali da «eliminare o ridurre il più possibile il rischio di output potenzialmente distorti che influenzano gli input per operazioni future» (c.d. feedback loops o circuiti di feedback) e, per tali circuiti, di «adeguate misure di attenuazione» (articolo 15, par. 4 AI Act);

- in relazione ai rischi di “alterazioni”, l'adozione di misure atte a garantire che i sistemi siano «resilienti ai tentativi di terzi non autorizzati di modificarne l'uso, gli output o le prestazioni sfruttando le vulnerabilità» degli stessi, in particolare attraverso «soluzioni tecniche volte a garantire la cibersecurity [...] adeguate alle circostanze e ai rischi pertinenti» (articolo 15, par. 5 AI Act).

Tra tali soluzioni tecniche sono comprese «misure volte a prevenire, accertare, rispondere, risolvere e controllare gli attacchi che cercano di manipolare il set di dati di addestramento» (c.d. data poisoning o «avvelenamento dei dati») o gli attacchi diretti verso «i componenti preaddestrati utilizzati nell'addestramento» (c.d. model poisoning o «avvelenamento dei modelli») ovvero «gli input progettati in modo da far sì che il modello di IA commetta un errore» (c.d. adversarial examples o «esempi antagonisti»), altresì definiti come model evasion o «evasione dal modello»), nonché «gli attacchi alla riservatezza o i difetti del modello»;

- l'adozione di un «sistema di gestione della qualità» atto a garantire la conformità del sistema di intelligenza artificiale ad alto rischio alla normativa regolatoria, articolato in «politiche, procedure e istruzioni scritte» e opportunamente «documentato in modo sistematico e ordinato» (articolo 17 AI Act);
- l'adozione, da parte dei provider, di «misure correttive necessarie per rendere conforme» il sistema di intelligenza artificiale ad alto rischio, «ritirarlo, disabilitarlo o richiamarlo, a seconda dei casi», allorché i fornitori ritengano o abbiano motivo di ritenere che il suddetto sistema «da essi immesso sul mercato o messo in servizio [...] non sia conforme», nonché l'adempimento di un dovere di informazione verso il distributore, il deployer, il rappresentante autorizzato e gli importatori (articolo 20, par. 1 AI Act);
- l'attuazione di «procedure di valutazione della conformità» (c.d. conformity assessment), in relazione ai soli sistemi di intelligenza artificiale ad alto rischio di cui all'Allegato III, punto 1 (ossia nel settore della biometria) per cui sono state applicate le norme armonizzate di cui all'articolo 40 AI Act o le specifiche comuni di cui all'articolo 41 AI Act, finalizzate a dimostrare la compliance del sistema ai requisiti previsti dalla Sezione II dell'AI Act (articolo 43 AI Act).

Un tema che, invece, resta aperto – e che impatta direttamente anche sulle prospettive di immediata applicabilità della fattispecie penale dell'articolo

437-bis, commi 1 e 3 nella parte riferita alle misure tecniche "di sicurezza" (essendo il comma 4 riferito esclusivamente agli obblighi di sorveglianza umana) – riguarda l'eventuale rilevanza di ulteriori normative, diverse dall'AI Act, in cui possano essere rinvenuti, in relazione ai sistemi di intelligenza artificiale ad alto rischio, in quanto tali o in quanto "prodotti" o "componenti di sicurezza" di prodotti, obblighi specifici inerenti all'adozione di misure tecniche di "sicurezza" con la funzione di prevenire malfunzionamenti o alterazioni del sistema.

Da questo punto di vista, ad esempio, deve essere evidenziato che già l'articolo 6 AI Act, nel contemplare le tipologie di sistemi di intelligenza artificiale ad alto rischio, prende in considerazione, oltre alle casistiche legati ai «settori» dell'Allegato III (par. 2), le ipotesi in cui il sistema è «*destinato a essere utilizzato come componente di sicurezza di un prodotto, o il sistema di IA è esso stesso un prodotto, disciplinato dalla normativa di armonizzazione dell'Unione elencata nell'allegato I*» (par. 1, lett. a)) ovvero i casi in cui «*il prodotto, il cui componente di sicurezza a norma della lettera a) è il sistema di IA, o il sistema di IA stesso in quanto prodotto, è soggetto a una valutazione della conformità da parte di terzi ai fini dell'immissione sul mercato o della messa in servizio di tale prodotto ai sensi della normativa di armonizzazione dell'Unione elencata nell'allegato I*».

Quanto sopra implica che un sistema di intelligenza artificiale riconducibile alle casistiche sopra descritte – in cui è «*componente di sicurezza*» o «*prodotto*» – allorché sia ricompreso nel campo di applicazione delle normative di armonizzazione dell'Unione europea di cui all'Allegato I, per ciò solo è qualificabile come ad alto rischio, rientrando per tale via nel cono applicativo dell'articolo 437-bis, comma 1 c.p., ed è al contempo soggetto agli eventuali e ulteriori requisiti e obblighi tecnici di "sicurezza" (si pensi, tra gli altri, a quelli previsti dal Regolamento (UE) 2023/1230 relativo alle macchine o al Regolamento (UE) 2016/425 sui dispositivi di protezione individuale) che siano previsti dalla normativa settoriale di riferimento richiamata nello stesso Allegato I per le fasi di sua progettazione, produzione, immissione sul mercato o anche addestramento.

Nella misura in cui quel sistema di intelligenza artificiale sia sottoposto, al di là delle previsioni generali dell'AI Act (allo stato non ancora applicabili), alla disciplina di settore rinvenibile *aliunde*, in relazione alla sua natura di «*prodotto*» o di «*componente di sicurezza*» di un prodotto, è chiaro che i suddetti obblighi tecnici di "sicurezza" – purché di essi si possa predicare che sono finalizzati a prevenire malfunzionamenti o alterazioni – siano fin d'ora cogenti nelle fasi di progettazione, produzione e immissione sul mercato (oltre che addestramento) e, pertanto, dovendosi ritenere le connesse «*misure tecniche di sicurezza*» già «*previste*» (ossia applicabili), l'eventuale omissione (totale o parziale) di esse sarà di immediata e attuale rilevanza

penale, anche prima delle nuove scadenze dettate dall'AI Act per i sistemi di intelligenza artificiale ad alto rischio.

Una questione per certi versi connessa e collaterale a quelle appena esaminate concerne la rilevanza da attribuire – quanto meno dal momento in cui saranno divenuti efficaci gli obblighi dell'AI Act per i sistemi di intelligenza artificiale ad alto rischio e che diverrà cogente per essi, nella propria veste di sistemi (e non di prodotti o componenti di sicurezza di prodotti), l'adozione delle «*misure tecniche di sicurezza*» previste dall'articolo 437-bis, comma 1 c.p. – a normative tecniche con funzione regolatoria preventiva diverse dall'AI Act, i cui requisiti di "sicurezza" possano ritenersi indirettamente richiamati dai pertinenti requisiti "diretti" previsti dall'AI Act. In tal senso, vengono in considerazione, ad esempio, tra i vari obblighi previsti dall'AI Act, la necessità che i sistemi di intelligenza artificiale ad alto rischio siano «*progettati e sviluppati in modo tale da conseguire un adeguato livello di accuratezza, robustezza e cibersecurity e da operare in modo coerente con tali aspetti durante tutto il loro ciclo di vita*» (articolo 15, par. 1 AI Act).

Dal punto di vista della *robustness* del sistema di intelligenza artificiale ad alto rischio, in particolare, risalta la doverosità di «*misure tecniche e organizzative*» volte a rendere i sistemi di intelligenza artificiale ad alto rischio «*il più resilienti possibile per quanto riguarda errori, guasti o incongruenze che possono verificarsi all'interno del sistema o nell'ambiente in cui esso opera, in particolare a causa della loro interazione con persone fisiche o altri sistemi*», nonché di «*soluzioni tecniche di ridondanza, che possono includere piani di backup o fail-safe*» (articolo 15, par. 4 AI Act); mentre, nella prospettiva della *cybersecurity*, l'esigenza che i sistemi di intelligenza artificiale ad alto rischio siano «*resilienti ai tentativi di terzi non autorizzati di modificarne l'uso, gli output o le prestazioni sfruttando le vulnerabilità del sistema*» attraverso «*soluzioni tecniche volte a garantire la cibersecurity [...] adeguate alle circostanze e ai rischi pertinenti*», ivi incluse quelle «*finalizzate ad affrontare le vulnerabilità specifiche dell'IA*» tra cui rilevano «*misure volte a prevenire, accertare, rispondere, risolvere e controllare gli attacchi che cercano di manipolare il set di dati di addestramento [...] o i componenti preaddestrati utilizzati nell'addestramento [...], gli input progettati in modo da far sì che il modello di IA commetta un errore [...] gli attacchi alla riservatezza o i difetti del modello*» (articolo 15, par. 5 AI Act).

Ebbene, ponendo mente ai richiamati requisiti dell'articolo 15 AI Act, il collegamento più immediato e abbastanza intuitivo che può svolgersi – anche in ragione dell'assonanza linguistica tra le terminologie utilizzate nei due diversi contesti – è riferito alle disposizioni della Direttiva (UE) 2022/2555 («*relativa a misure per un livello comune elevato di cibersecurity nell'Unione*», meglio noto come NIS2) e al relativo atto di recepimento nazionale rappresentato dal d.lgs. 138/2024, il quale – si rammenta – tra i vari

obblighi previsti, impone agli organi di amministrazione e agli organi direttivi dei soggetti essenziali e dei soggetti importanti di approvare «*le modalità di implementazione delle misure di gestione dei rischi per la sicurezza informatica adottate*» (articolo 23, comma 1, lett. a)), intese quali «*misure tecniche, operative e organizzative adeguate e proporzionate [...] alla gestione dei rischi posti alla sicurezza dei sistemi informativi e di rete che tali soggetti utilizzano nelle loro attività o nella fornitura dei loro servizi, nonché per prevenire o ridurre al minimo l'impatto degli incidenti per i destinatari dei loro servizi e per altri servizi*» (articolo 24, comma 1), atte a garantire una serie di elementi, tra cui «*gestione degli incidenti*», «*continuità operativa, ivi inclusa la gestione di backup, il ripristino in caso di disastro, ove applicabile, e gestione delle crisi*», nonché «*sicurezza dell'acquisizione, dello sviluppo e della manutenzione dei sistemi informativi e di rete, ivi comprese la gestione e la divulgazione delle vulnerabilità*».

Allo stesso modo, potrebbe essere chiamato in causa il comparto normativo del Regolamento (UE) 2022/2554 («*relativo alle resilienza operativa digitale per il settore finanziario*», meglio noto come *Digital Operational Resilience Act*, DORA) e delle relative disposizioni di adeguamento contenute nel d.lgs. 23/2025, il quale prevede che «*le entità*

finanziarie monitorano e controllano costantemente la sicurezza e il funzionamento dei sistemi e degli strumenti di TIC e riducono al minimo l'impatto dei rischi informatici sui sistemi di TIC adottando politiche, procedure e strumenti adeguati per la sicurezza delle TIC», anche attraverso l'adozione e l'attuazione di «*politiche, procedure, protocolli e strumenti per la sicurezza delle TIC miranti a garantire la resilienza, la continuità e la disponibilità dei sistemi di TIC, in particolare quelli a supporto di funzioni essenziali o importanti, nonché a mantenere standard elevati di disponibilità, autenticità, integrità e riservatezza dei dati conservati, in uso o in transito*», nonché di «*soluzioni e processi TIC*» che «*garantiscono la sicurezza dei mezzi di trasferimento dei dati*», «*riducono al minimo i rischi di corruzione o perdita di dati, di accesso non autorizzato nonché di difetti tecnici che possono ostacolare l'attività commerciale*», «*prevengono la mancanza di disponibilità, il deterioramento dell'autenticità o dell'integrità, le violazioni della riservatezza e la perdita di dati*» e «*assicurano la protezione dei dati contro i rischi derivanti dalla gestione dei dati, compresi la cattiva amministrazione, i rischi relativi al trattamento dei dati e l'errore umano*» (articolo 9, parr. 1, 2 e 3).

Al di là di tali rinvii esemplificativi, in questa prospettiva – che gli operatori della filiera dell'intelligenza artificiale dovranno tenere in dovuta considerazione, anche in sede preliminare di analisi dei rischi interni – l'AI Act sembra poter rappresentare una normativa “aperta” verso l'esterno, capace di attrarre nella propria orbita applicativa – e per conseguenza nel perimetro punitivo dell'articolo 437-bis, comma 1 c.p. – l'osservanza di ulteriori ed eterogenei fasci di obblighi di sicurezza “derivati” o secondari, che possano

ritenersi richiamati in via indiretta o comunque desumibili in via interpretativa dai vari requisiti "originari" o "primari" posti dal regolamento medesimo.

Passando all'analisi della seconda condotta incriminata nell'articolo 437-bis, comma 1 c.p. – relativa alla omessa adozione di «*misure di sorveglianza umana*» – si rinvia, per ragioni di sintesi espositiva, all'analisi svolta nel precedente paragrafo 2 in materia di obblighi di *human oversight*, che forgiavano l'area di rilevanza penale disegnata dalla fattispecie.

In questa sede, tuttavia, occorre mettere in evidenza un tratto di relevantissima novità – dal punto di vista della definizione del fatto tipico – rispetto alle originarie previsioni dello schema di decreto legislativo, dal momento che l'omissione delle misure di sorveglianza umana, già incriminata nel comma 1, rappresenta oggi l'unica condotta penalmente rilevante per la posizione del *deployer*, i cui profili di responsabilità sono ora stati autonomamente considerati nel "nuovo" comma 4, introdotto soltanto nella versione finale del d.lgs. 160/2026, all'insegna di una netta differenziazione di trattamento rispetto ai soggetti attivi punibili ai sensi del comma 1.

Ciò significa che l'«*utilizzatore professionale*», invece, non sarà penalmente perseguibile per omissioni riferite all'adozione delle misure tecniche "di sicurezza" descritte nel comma 1, salvo il caso in cui – come in precedenza ipotizzato – egli sia partecipe della fase di «*addestramento*» del sistema di intelligenza artificiale ad alto rischio, nel cui ambito potranno essere poste a suo carico anche eventuali omissioni inerenti alle misure tecniche suddette.

Sotto questo profilo, la condizione di "asimmetria" normativa che, nel contesto dell'articolo 437-bis c.p., caratterizza il *deployer*, a una prima analisi, pare giustificabile in considerazione del fatto che l'AI Act alloca la gran parte degli obblighi afferenti all'adozione di misure di "sicurezza" per i sistemi di intelligenza artificiale ad alto rischio in capo a soggetti diversi e, in particolare, al *provider*, laddove per la posizione del *deployer* un simile obbligo sembra principalmente individuabile soltanto nella disposizione dell'articolo 26, par. 1 AI Act, in forza della quale egli è tenuto ad adottare «*idonee misure tecniche e organizzative per garantire di utilizzare [i sistemi di intelligenza artificiale ad alto rischio] conformemente alle istruzioni per l'uso che accompagnano i sistemi*».

Ponderata o meno che sia stata la scelta del legislatore delegato, è evidente che un simile "ritaglio" dell'area di rilevanza penale riservata, per il tramite del nuovo comma 4, alle condotte riferibili all'«*utilizzatore professionale*» alleggerisce notevolmente il peso delle responsabilità in capo al *deployer*, restringendole alle ipotesi di mancata o inadeguata adozione di misure di *human oversight* sui sistemi di intelligenza artificiale ad alto rischio, tra l'altro – come in seguito si osserverà – soltanto in casistiche caratterizzate da un elemento soggettivo di particolare pregnanza, con esclusione di omissioni meramente colpose.

L'opzione in favore di un'incriminazione minima della posizione del *deployer*, all'evidenza, risponde all'intento di contenere la proliferazione delle responsabilità penali in capo ai soggetti della filiera dell'intelligenza artificiale – comunque pur sempre assoggettati, ivi incluso il *deployer* stesso, alle impattanti sanzioni amministrative dal sapore autenticamente punitivo – in considerazione della vastità della platea dei soggetti che, sul mercato, sono destinati a rivestire il ruolo di «*utilizzatore professionale*» di un sistema di intelligenza artificiale ad alto rischio.

Proseguendo l'esame della fattispecie del comma 1, si osserva che l'anticipazione della tutela penale allo stadio del pericolo – scelta che si pone in diretta relazione con il rango delle oggettività giuridiche potenzialmente coinvolte dalla condotta illecita – trovava, almeno nello schema del decreto legislativo, una parziale moderazione nella configurazione dell'offesa nella forma del pericolo concreto, poiché, in tale versione del reato, la condotta superava la soglia di rilevanza penale soltanto «*quando da tali omissioni derivi pericolo concreto per la vita o l'incolumità pubblica o individuale*».

La versione finale dell'articolo 437-bis, comma 1 c.p., contenuta nel d.lgs. 160/2026, abbandona il modello del reato di pericolo concreto, per il tramite della elisione del suddetto requisito all'interno della fattispecie, che ora assume rilevanza penale «*quando da tali omissioni derivi pericolo per la vita o l'incolumità pubblica o individuale*».

Tale modifica del testo normativo rispetto allo schema di decreto legislativo – che, come si vedrà, riguarda anche la fattispecie del secondo periodo – determina rilevanti ricadute anche e soprattutto dal punto di vista processuale, poiché il passaggio da una fattispecie articolata su una situazione espressamente qualificata in termini di «*pericolo concreto*» per i beni giuridici menzionati (che impone all'accusa e al giudice penale uno sforzo, rispettivamente, probatorio e motivazionale senz'altro superiore) a una fattispecie di mero «*pericolo*» (non ulteriormente qualificato) potrebbe aprire uno spiraglio a letture indebitamente estensive circa la natura astratta del requisito di pericolo, il quale, com'è noto, implica una notevole semplificazione probatoria, ponendo la difesa in una posizione di partenza di "armi spuntate".

In ogni caso, la clausola in esame sarà presumibilmente oggetto del futuro dibattito, giurisprudenziale e dottrinale, in relazione alla funzione dogmatica che essa è destinata a svolgere nel contesto della fattispecie incriminatrice, con specifico riferimento alla categoria della punibilità.

In assenza di un concreto supporto ricostruttivo fornito dalla Relazione illustrativa allo schema di decreto legislativo (che genericamente osserva come «*[l]a punibilità è subordinata al verificarsi di un pericolo concreto*»), sarà necessario chiarire, invero, se essa debba essere interpretata come elemento costitutivo del fatto tipico, nella specie dell'offesa (dunque, nell'accezione di

una condizione di punibilità “intrinseca”), ovvero come elemento estraneo alla tipicità del fatto, quale condizione “obiettiva” di punibilità.

L’opzione ermeneutica prescelta non è affatto indifferente, dal momento che – com’è noto – il regime delle condizioni obiettive di punibilità (articolo 44 c.p.) rende la circostanza oggetto della condizione “insensibile” e indifferente rispetto alla volontà dell’agente, con la conseguenza che l’evento (offensivo) gli sarà imputato sulla base del mero riscontro di una relazione di causalità materiale con la condotta.

Nella versione della disposizione contenuta nello schema di decreto legislativo, il legislatore delegato sembrava essere incorso in un *lapsus calami*, giacché il riferimento al bene giuridico della «*incolumità pubblica*» compariva per ben due volte, tanto nel primo periodo (accostato all’incolumità «*individuale*») quanto nel secondo periodo della disposizione (giustapposto, come si vedrà, alla «*sicurezza dello Stato*»).

La versione definitiva dell’articolo 437-*bis*, comma 1 c.p., invero, rettifica la svista contenuta nel testo, confermando l’intenzione del legislatore delegato di inaugurare – rispettivamente, nel primo e nel secondo periodo del comma 1 – una “scalarità” offensiva tra fattispecie, la quale, tuttavia, diversamente da come ci si sarebbe attesi, non è costruita in termini di progressione qualitativa dai beni individuali (vita e della incolumità fisica) verso i beni super-individuali e diffusi (incolumità pubblica e sicurezza dello Stato), bensì è attuata mediante aggregazione “tipologica” delle oggettività giuridiche tutelate, che ora figurano in serie all’interno della medesima disposizione incriminatrice del primo periodo («*la vita o l’incolumità pubblica o individuale*»).

Dal punto di vista sanzionatorio, la cornice edittale per l’ipotesi base del comma 1, nonché per la correlata fattispecie autonoma riservata al *deployer* nel comma 4 – la reclusione da uno a cinque anni – rende i delitti astrattamente “permeabili” rispetto all’applicazione di vari istituti di diritto sostanziale e processuale, vuoi *in bonam parte* vuoi *in malam partem*, calibrati sui minimi e/o sui massimi di pena, quali, ad esempio, la non punibilità per particolare tenuità del fatto (articolo 131-*bis* c.p.), le misure cautelari personali, coercitive (inclusa la custodia cautelare in carcere) e interdittive, nonché – per le ipotesi (secondo periodo) che rientrano nei *range* edittali previsti (reclusione superiore nel massimo a cinque anni) – le intercettazioni telefoniche, anche tra presenti, nonché le intercettazioni telematiche.

Il comma 1, nel secondo periodo, prevede poi una seconda *figura criminis*, che ricorre – come anticipato – quando l’offesa compie un “salto di qualità”, nel senso che dalla condotta sia derivato «*un pericolo per la sicurezza dello Stato*», che giustifica una notevole dilatazione (soprattutto nel massimo) della forbice edittale, compresa tra due e otto anni di reclusione, all’insegna di un trattamento sanzionatorio ancora una volta parificato tanto per i soggetti

punibili ai sensi del comma 1 che per il *deployer* punibile ai sensi dell'omologa (seppur non identica) fattispecie del comma 4.

Al netto della divaricazione linguistica presente nel secondo periodo – in cui è individuato come fattore da cui deve scaturire il pericolo il «*fatto*» e non, come nel primo periodo, le «*omissioni*» (per cui *infra*) – spetterà agli interpreti chiarire, anzitutto, la qualificazione giuridica della fattispecie (circostanza aggravante o fattispecie autonoma), benché la *littera legis*, ponendo in apertura di disposizione un rinvio integrale al «*fatto*» descritto nel primo periodo del comma 1, fornisca un indizio che, seppur non determinante, potrebbe deporre, in applicazione del criterio nominalistico, in favore della natura circostanziale della fattispecie (la quale, peraltro, così qualificata, sarebbe suscettibile di bilanciamento eterogeneo con le circostanze attenuanti concorrenti ai sensi dell'articolo 69 c.p., senza il "contrappeso" di una disposizione normativa *ad hoc* che ne moduli la portata).

L'accennata diversificazione terminologica presente nei due periodi della disposizione, tra l'altro, rende assai incerta e discutibile la ricostruzione della fattispecie dal punto di vista dell'offensività, giacché le disposizioni incriminatrici del primo e del secondo periodo non appaiono linguisticamente "allineate": in particolare, la situazione di pericolo (per la sicurezza dello Stato) del secondo periodo – almeno così il dato normativo induce a ritenere – non appare correlata direttamente alle «*omissioni*» del primo periodo, bensì al «*fatto*» ivi descritto.

Il riferimento posto in apertura del secondo periodo al «*fatto*» (e non alle «*omissioni*») legittima, pur se non certifica, una ricostruzione del secondo periodo come fattispecie avente natura "complessa" e struttura causalmente "concatenata", destinata a ricorrere nel caso in cui una condotta consistente nell'omissione di misure tecniche di "sicurezza" e/o di sorveglianza umana abbia prodotto una situazione di pericolo per la vita o l'incolumità (pubblica o individuale) – sempre ammesso e non concesso che il pericolo costituisca elemento costitutivo del fatto – dalla quale, a propria volta, sia conseguita una situazione di pericolo per la sicurezza dello Stato.

Così interpretata, tuttavia, la fattispecie del secondo periodo perderebbe, proprio dal punto di vista dell'offensività, la propria natura "alternativa" e "parallela" rispetto a quella del primo periodo, in ragione della quale, secondo una lettura più piana delle due disposizioni, ove dalle omissioni descritte nel primo periodo scaturisca un pericolo per i beni "personali", il fatto sarà punibile ai sensi di tale disposizione, mentre qualora dalle medesime omissioni discenda un pericolo per il bene sovraindividuale della sicurezza dello Stato, il fatto sarà punibile ai sensi del secondo periodo.

L'utilizzo della parola «*fatto*» – probabilmente ritenuta neutra e intercambiabile (con quella di "omissione") da parte del legislatore delegato – innesci invece una serie di criticità interpretative che rischiano di annichilire, ancora prima che essa entri a pieno regime, l'applicazione della

fattispecie del secondo periodo, rendendola, se non di difficile e residuale riscontro pratico, quanto meno estremamente "ostica" in termini di accertamento probatorio secondo i canoni propri del processo penale.

Un cortocircuito interpretativo paragonabile a quello generato dall'articolo 437-bis, comma 1, secondo periodo c.p. sembra potersi rinvenire – nella diversa *sedes materiae* dei delitti contro l'ambiente (Titolo VI-bis del Codice penale, come introdotto dalla l. 68/2015) – nell'articolo 452-*quinqües* c.p., il quale, nel porre la disciplina dei «[d]elitti colposi contro l'ambiente», prevede, nel comma 1, le fattispecie colpose (di evento) di inquinamento ambientale e di disastro ambientale e, nel comma 2, una fattispecie di "pericolo" che ricorre quando «*dalla commissione dei fatti di cui al comma precedente deriva il pericolo di inquinamento ambientale o di disastro ambientale*».

La locuzione posta dall'articolo 452-*quinqües*, comma 2 c.p. – analoga a quella dell'articolo 437-bis, comma 1, secondo periodo c.p. – ha parimenti posto in letteratura il tema del significato da attribuire alla parola «*fatti*», giacché intendendo il termine in senso tecnico si sarebbe giunti all'autentico paradosso di ipotizzare una fattispecie in cui da un fatto (comprensivo di evento) di inquinamento ambientale o di disastro ambientale già perfezionatosi derivi un mero «*pericolo di inquinamento ambientale o di disastro ambientale*».

Il "vicolo cieco" a cui simile interpretazione stretta del dato normativo avrebbe condotto ha ben presto portato gli interpreti, in tale contesto di disciplina, a rileggere il termine «*fatti*» alla stregua di "condotte", in modo da rendere coerente e così concretamente applicabile la fattispecie di pericolo dell'articolo 452-*quinqües*, comma 2 c.p.

La stessa soluzione interpretativa potrebbe essere riproposta, per uscire dal *cul-de-sac* ermeneutico, anche per la fattispecie (autonoma o circostanziale che sia) dell'articolo 437-bis, comma 1, secondo periodo c.p., leggendo la disposizione incriminatrice come se prevedesse il descritto incremento sanzionatorio qualora «*dalle omissioni del comma 1 derivi un pericolo per la sicurezza dello Stato*».

Per ultimo, si osserva, sempre dal punto di vista dell'offensività, che rispetto alle fattispecie del comma 1 (e del comma 4) il legislatore delegato non ha previsto delle corrispondenti ipotesi (autonome o aggravate, che dir si voglia) di evento di danno, replicando il paradigma di incriminazione già seguito per altre fattispecie di "disastro" contenute all'interno del Titolo VI del Codice penale, quali, tra le altre, il disastro innominato (articolo 434, comma 2 c.p.) e la rimozione od omissione di cautele antinfortunistiche (articolo 437, comma 2 c.p., che per espressa affermazione della Relazione illustrativa allo schema di decreto legislativo rappresenta il prototipo dell'incriminazione in esame).

La scelta si spiega, forse, con la difficoltà di ipotizzare e, quindi, di tipizzare (rendendolo suscettivo di prova) un evento – fenomenologicamente

sussequente alla situazione di pericolo concreto per i beni individuali e sovraindividuali richiamati nel primo e nel secondo periodo – di disastro “da intelligenza artificiale”, da colpire con un incremento sanzionatorio frazionario rispetto alla fattispecie base e aggravata.

Il delitto previsto dal comma 1, dal punto di vista dell’elemento soggettivo, è punito a titolo di dolo generico, in termini di rappresentazione e di volontà, che devono investire tutti gli elementi costitutivi del fatto tipico, incluso (ove non interpretato come condizione obiettiva di punibilità) il requisito del pericolo concreto per i beni menzionati.

Quanto alla componente volontaristica del dolo, a una prima analisi, la fattispecie risulta compatibile con tutte le varie graduazioni dell’elemento volitivo, comprensive sia dolo intenzionale che del dolo diretto (quando l’evento offensivo è previsto come conseguenza certa o altamente probabile della condotta, quale evento accessorio e collaterale al risultato perseguito), nonché con una possibile contestazione a titolo di dolo eventuale, per quelle ipotesi in cui si possa ritenere – con il conforto e il supporto probatorio di uno o più degli “indicatori” elaborati dalla giurisprudenza (tra le altre, dalla nota sentenza delle Sezioni Unite nel caso *ThyssenKrupp*) – che il/i soggetto/i della catena del valore dell’intelligenza artificiale, mutuando i principi espressi dalle stesse Sezioni Unite, abbia avuto una «chiara, lucida, rappresentazione» dell’evento offensivo in termini di «*apprezzabile livello di probabilità*» e lo abbia accettato, attraverso una «*presa di posizione volontaristica*» in guisa di «*adesione all’evento, che consenta di scorgervi un atteggiamento ragionevolmente assimilabile alla volontà*».

Ai fini dell’accertamento del dolo eventuale del soggetto agente, verranno in considerazione i “fattori” di natura indiziaria affermati dalla pronuncia nel caso *ThyssenKrupp* – i quali rapportati al contesto applicativo, tipicamente aziendale, dell’intelligenza artificiale assumono particolare pregnanza – quali la «*condotta che caratterizza l’illecito*» (in relazione alle sue modalità e ai suoi mezzi), la «*lontananza dalla condotta standard*» (rilevante soprattutto «*negli ambiti governati da discipline cautelari*»), la «*storia e le precedenti esperienze*» (anche in relazione a precedenti condotte illecite specifiche) e la «*personalità*» dell’agente (correlata a «*la sua cultura, l’intelligenza, la conoscenza del contesto in cui sono maturati i fatti*»), la «*durata e la ripetizione della condotta*», la «*condotta successiva al fatto*» (che consente di valorizzare anche eventuali contro-condotte riparatorie realizzate nell’immediatezza del fatto), la «*probabilità di verifica dell’evento*» e il «*contesto lecito o illecito*» in cui si è svolto il fatto (nel senso che «*una situazione illecita di base indizia più gravemente il dolo*»).

Sotto il profilo dell’elemento psichico del reato, deve essere posta in evidenza una differenziazione altrettanto rilevante introdotta dal d.lgs. 160/2026, rispetto alla versione originaria contenuta nello schema di decreto legislativo, nel comma 4, che configura – come già detto – una responsabilità autonoma

del *deployer*, il quale tuttavia deve omettere «*intenzionalmente*» le misure di sorveglianza umana.

Ciò implica che, rispetto alla figura del *deployer*, il perimetro della fattispecie penale è senz'altro più ristretto non solo dal punto di vista oggettivo (rilevando la sola condotta di omissione di misure di *human oversight*), bensì anche dal punto di vista soggettivo, potendo egli rispondere del reato del comma 4 soltanto a titolo di dolo intenzionale, il quale – com'è noto – impone un accertamento particolarmente pregnante della componente psichica del soggetto agente.

Si noti, peraltro, che, a fronte di simile differenziazione del coefficiente soggettivo (più significativo) richiesto in capo al *deployer*, il trattamento sanzionatorio a lui destinato è in tutto e per tutto – rispetto a entrambe le situazioni di pericolo (per i beni “personali” e per sicurezza dello Stato) che abbia determinato con la propria condotta – parificato, almeno a livello di comminatoria edittale astratta, alle pene previste per i fatti del comma 1, punibili (anche) a mero titolo di dolo generico, circostanza suscettibile di ingenerare un profilo di disuguaglianza-irragionevolezza (articoli 3 e 27, comma 1 Cost.) censurabile in sede costituzionale.

Dal punto di vista della durata, infine, le fattispecie del comma 1 (e del comma 4) si configura come reato permanente, in quanto tale caratterizzato dalla presenza di una fase successiva al suo perfezionamento – che avviene con il mancato compimento dell'azione giuridicamente doverosa tipizzata dal legislatore (nel caso di specie, l'adozione delle misure tecniche di sicurezza e/o di sorveglianza umana, come detto, cogenti soltanto a partire dalle scadenze temporali dettate dall'AI Act per i sistemi di intelligenza artificiale ad alto rischio) – e di durata non preventivamente determinabile, a cui è sottesa una condotta continuativa e volontaria dell'agente che determina il protrarsi dell'offesa punibile al/i bene/i giuridici tutelato/i.

La natura permanente del reato – com'è intuibile – riverbera effetti sotto il profilo sostanziale e processuale, con particolare riferimento alla procrastinazione della decorrenza del termine di prescrizione del reato a partire dal momento della sua cessazione (articolo 158, comma 1 c.p.).

La permanenza, dunque, protrae la fase di consumazione del reato e cessa soltanto nel momento in cui il soggetto agente pone in essere l'azione doverosa richiesta dall'ordinamento, ossia, nel caso in esame, adottando le misure di “sicurezza” (tecniche e/o di sorveglianza umana) la cui omissione costituisce il fulcro dell'incriminazione.

Sul punto, tuttavia, con riferimento al reato del comma 1, sembra riproporsi la problematica in precedenza esaminata circa il significato da attribuire, nel contesto della fattispecie penale, al requisito di “adeguatezza” delle misure, giacché, intendendolo quale sinonimo di “idoneità” rispetto al fine preventivo indicato dalla disposizione («*prevenire malfunzionamenti o alterazioni del funzionamento dei sistemi*»), potrebbe ritenersi non sufficiente, ai fini della

cessazione della permanenza, l'adozione o (ove le misure fossero presenti in parte) l'adeguamento di qualsivoglia misura tecnica di sicurezza.

Così opinando, infatti, avrebbe efficacia realmente interruttiva della consumazione *in fieri* del reato la sola implementazione di misure che possano essere considerate «*adequate*», con tutte le incognite già illustrate circa il parametro di riferimento per giudicare della loro «adeguatezza-idoneità» e, soprattutto, nella totale incertezza circa il fatto che le misure eventualmente adottate dall'agente, soltanto *ex post*, in sede giudiziale, siano considerato in tutto o in parte non «*adequate*».

In tale chiave di lettura, il soggetto agente – e, così, la sua organizzazione di appartenenza – potranno maturare una ragionevole aspettativa di «adeguatezza-idoneità» delle misure adottate, nella prospettiva di cessazione della permanenza della condotta omissiva, forse soltanto nell'ipotesi di avvenuta irrogazione da parte delle autorità regolatorie competenti, in relazione alla/e violazione/i di rilevanza amministrativa sottesa/e al caso di specie, della misura sanzionatoria non pecuniaria «in forma specifica» consistente nell'ordine di «*eliminare le infrazioni e di astenersi dal ripeterle*», nella sua variante prescrittiva che indichi «*le misure da adottare e il termine per attuarle*» (articolo 22, comma 1, lett. a) d.lgs. 179/2026).

In tale ipotesi, potrebbe ritenersi – sulla falsariga di quanto nella prassi accade nell'ambito di meccanismi sanzionatori e/o estintivi che prevedono una «regolarizzazione» postuma «su impulso» dell'organo di vigilanza (si pensi alle oblazioni «amministrative» ben note nel settore della salute e sicurezza del lavoro, dell'ambiente e degli alimenti o, ancora, ai «*poteri correttivi*» di natura ingiunzionale previsti nel contiguo ambito sanzionatorio della *privacy*, ai sensi dell'articolo 58, par. 2 GDPR) – che l'adempimento dell'ordine-sanzione irrogato dall'autorità competente, attraverso la realizzazione, nel termine indicato, delle misure che quest'ultima ha ritenuto conferenti al caso di specie ai fini della cessazione della violazione (intesa come violazione della disciplina regolatoria dell'AI Act, di rilevanza amministrativa, implicitamente «assorbita» dal precetto penale), fornisca «garanzia» circa il carattere di «adeguatezza» delle misure apportate.

A ben vedere, il caso illustrato appare più di scuola che reale, dal momento che simile misura «in forma specifica», come tutte le altre misure sanzionatorie non pecuniarie previste dall'articolo 22 d.lgs. 179/2026, può trovare applicazione soltanto a fronte di «*violazioni del regolamento IA caratterizzate da scarsa offensività o pericolosità*», dunque in situazioni di dubbia compatibilità con fatti che – nella adiacente sede penale – hanno determinato un pericolo concreto per beni individuali e sovraindividuali, quali la vita, l'incolumità e la sicurezza dello Stato.

Le condotte omissive incriminate nel comma 1 assumono rilevanza penale – ai sensi della previsione espressa del comma 3 (da mettere in relazione

all'articolo 42, comma 2 c.p.) – anche qualora siano psicologicamente connotate da colpa.

La fattispecie colposa in esame sarà senz'altro destinata, nella prassi, a maggior successo applicativo rispetto alle corrispondenti ipotesi dolose, in considerazione del naturale "ambiente" – quello delle organizzazioni complesse che svolgono attività "lecite di base" – in cui può trovare concretizzazione il *deficit* di misure preventive (tecniche di sicurezza o di sorveglianza umana) sanzionato da legislatore delegato.

Infatti, è il contesto di impresa dei fornitori, degli importatori, dei rappresentanti autorizzati e degli utilizzatori professionali (nei limiti in cui svolgano attività di «*addestramento*») – che chiama spesso in causa le realtà dei grandi *player* del mercato dell'intelligenza artificiale – l'ambito fisiologico nel quale assumeranno prioritaria rilevanza eventuali violazioni non certo consapevolmente e intenzionalmente realizzate, bensì addebitabili a titolo di colpa generica o di colpa specifica (in quanto inosservanti delle disposizioni dell'AI Act, della normativa interna di adeguamento e delle eventuali norme tecniche di settore).

Il rimprovero per il fatto colposo, tuttavia, è graduato dal legislatore delegato – nella già menzionata prospettiva del diritto penale "minimo" – nel senso che l'addebito ai sensi dell'articolo 437-*bis*, comma 3 c.p. si deve attestare su di un livello «*colpa grave*», la quale sola – dunque, con esclusione della colpa "semplice" – rende l'omissione punibile con pena «*ridotta da un terzo a un sesto*».

In questa direzione, secondo le indicazioni della Relazione illustrativa allo schema di decreto legislativo, «*[l]a selezione della colpa grave come elemento di tipicità – sul modello dell'art. 590-sexies c.p. introdotto dalla legge n. 24 del 2017 – risponde a un'esigenza di proporzionalità: si circoscrive la punibilità alle macroscopiche violazioni delle regole di diligenza, evitando la criminalizzazione di ogni scostamento dallo standard tecnico in un dominio caratterizzato da rapida evoluzione e da una compresenza di standard tecnici nazionali, unionali e internazionali*».

In tal senso, dunque, il limite interno alla fattispecie penale della «*colpa grave*» implica che potranno essere colpite dallo stigma della sanzione penale – benché, si intende, determinino le situazioni di pericolo concreto poc'anzi descritte – non tanto le mere "disattenzioni" nell'analisi, nella valutazione e nell'implementazione delle misure tecniche di sicurezza da adottare (che potranno assumere rilevanza dal punto di vista delle pur impattanti sanzioni amministrative applicate dalle autorità di settore), quanto, piuttosto, eventuali scelte "disinvolte" da parte delle organizzazioni nella modulazione dei presidi e dei controlli per i sistemi di intelligenza artificiale ad alto rischio.

Sotto quest'ultimo profilo, ad esempio, in relazione a requisiti di sicurezza del sistema che rivestono particolare importanza ("core", quali quelli finalizzati a

contrastare i fenomeni del *data poisoning*, del *model poisoning* o del *model evasion*), uno scostamento "sostanziale" dallo *standard* di riferimento, in assenza di una motivazione ragionevole che abbia una "tenuta" dal punto di vista tecnico e giuridico, ovvero l'opzione per un livello di misure di sicurezza significativamente inferiori o, comunque, in concreto palesemente non adeguate alle criticità da fronteggiare, *ex post*, potrebbero essere ricondotte dal giudice penale vuoi a un contegno di "accettazione del rischio" (espressione di dolo eventuale) vuoi a una "sopravalutazione" delle capacità dell'organizzazione di scongiurare la violazione (espressione di colpa cosciente) e ciò – si badi bene – indipendentemente dal fatto che la violazione assuma rilevanza sanzionatoria o sia effettivamente sanzionata in sede amministrativa.

Al netto del quesito sulla natura della fattispecie del comma 3 (reato autonomo o circostanza attenuante), che dovrà essere risolta alla stregua dei criteri chiarificatori (ma non sempre dirimenti) elaborati dalla consolidata giurisprudenza di legittimità, la disposizione non codifica – come forse sarebbe lecito attendersi, stante il particolare tecnicismo della materia – specifici "indici" da quali dedurre (o in forza dei quali escludere) la gravità della colpa.

Sotto tale profilo, il legislatore delegato avrebbe potuto trovare un utile riferimento – almeno a livello di impostazione metodologica – nella recente proposta di articolato di riforma in materia di salute e sicurezza sui luoghi di lavoro, elaborata dalla Commissione di studio istituita presso il Ministero della Giustizia con d.m. 27 marzo 2024 (Commissione Sisto), in cui, nell'ambito di un inedito articolo 590-*septies* c.p. – che limita la punibilità del datore di lavoro a titolo di colpa grave, nel caso di adozione di un modello di organizzazione, gestione e controllo (conforme ai requisiti di idoneità previsti dall'articolo 30 d.lgs. 81/2008) – da un lato, sono espressamente tipizzati alcuni indici («*fattori*») che il giudice è tenuto a considerare, ai fini della valutazione sul grado della colpa, per escluderne la gravità (comma 2)⁹ e, dall'altro, sono individuati alcuni obblighi la cui violazione implica automaticamente un rimprovero per colpa grave (comma 3).

Detta disposizione avrebbe potuto costituire un modello di disciplina replicabile, *mutatis mutandis*, anche nel contesto regolatorio e sanzionatorio dell'intelligenza artificiale – connotato da una vorticoso evoluzione delle norme tecniche di settore (ancorché talora non del tutto emanate), da una tutt'oggi non piena conoscenza dei fattori di rischio e delle modalità

⁹ I «*fattori*» sono individuati nella «*natura e della complessità dell'attività*», nelle «*conoscenze sul rischio*», nelle «*buone prassi*» (in materia di tutela della salute e sicurezza sul lavoro validate dalla Commissione consultiva permanente per la salute e sicurezza sul lavoro), nel «*possesso della asseverazione*» del modello di organizzazione gestione e controllo (ai sensi dell'articolo 51, comma 3-*bis* d.lgs. 81/2008).

comportamentali delle *machinae* – quale parametro per graduare la colpevolezza dell'agente, nel rispetto del fondamentale canone di colpevolezza-personalità della responsabilità penale (articolo 27, comma 1 Cost.).

Occorre mettere in risalto, ancora una volta, un significativo profilo di differenziazione normativa del trattamento riservato al *deployer* nell'articolo 437-bis c.p. da parte del d.lgs. 160/2026, il quale, rispetto all'autonoma fattispecie di reato del comma 4, non ha reso penalmente rilevanti anche eventuali condotte commesse con colpa, dal momento che la disposizione del comma 3 richiama soltanto i «*fatti previsti dal comma primo*» e non essendo a tal fine sufficiente allo scopo il rinvio al comma 1, da intendersi solo *quoad poenam*, contenuto nel comma 4.

Consapevolmente o meno perseguita tale opzione da parte del legislatore delegato, è evidente, in ogni caso, che la limitazione della responsabilità penale del *deployer* per i soli fatti commessi a titolo di dolo (per di più intenzionale) e l'esclusione di un rimprovero a suo carico a titolo di colpa (ancorché soltanto grave) costituisce un trattamento sicuro "favore", che solleva l'ampia platea dei soggetti che utilizzano sistemi di intelligenza artificiale nell'ambito di un'attività «*professionale*» e «*non personale*» dallo "spauracchio" di contestazioni, di una certa gravità, in sede penale connesse a inadempimenti non voluti e/o a erronee valutazioni tecniche e/o a inefficienze dell'organizzazione e dei presidi interni in tema di misure di sorveglianza umana.

4.2. L'alterazione di sistemi di intelligenza artificiale ad alto rischio (articolo 437-bis, comma 2, c.p.).

La fattispecie incriminatrice del comma 2 si caratterizza rispetto a quella base del comma 1, per natura, materialità e disvalore del tutto eterogenei.

A livello di materialità, il reato, pur costituendo parimenti ipotesi delittuosa, è configurato come reato comune, avente natura commissiva, in quanto imperniato su una condotta di "alterazione" (sempre e soltanto) di sistemi di intelligenza artificiale ad alto rischio.

La natura comune del reato risulta pacifica, stante la piena compatibilità della condotta di "alterazione" sia con l'attività degli operatori della filiera dell'intelligenza artificiale sia – per quanto di maggior rilievo in questa sede – di soggetti "terzi" estranei, non direttamente destinatari degli obblighi posti dalla disciplina regolatoria di settore.

Tale lettura risulta coerente con le indicazioni contenute nella Relazione illustrativa allo schema di decreto legislativo, che, in modo ancora più *tranchant* sul punto dell'individuazione del soggetto agente, ritiene la disposizione incriminatrice preordinata *tout court* a «*colpire la condotta di chi, non avendo il controllo lecito del sistema di intelligenza artificiale, altera il funzionamento del sistema*», quale «*previsione di chiusura, necessaria per*

evitare aree di impunità nei casi in cui il pericolo concreto derivi dalla manipolazione attiva, esterna, del sistema».

Senonché, a fronte di un'*intentio legis* indiscutibilmente orientata a punire (solo ed esclusivamente) il fatto del "terzo" – esterno rispetto al "contesto" cui pertiene il sistema di intelligenza artificiale – il percorso di identificazione delle soggettività attive del reato richiede un ulteriore passaggio interpretativo, imposto dall'inserimento nel comma 2 della clausola *«al di fuori dei casi indicati al comma 1»*.

Si tratta, in particolare, di valutare se soltanto il "terzo" estraneo possa essere in effetti ricompreso – come intende il legislatore delegato – nel novero dei soggetti attivi o se, invece, il tenore della richiamata clausola, per eterogenesi dei fini, consentirà di ricondurre in via interpretativa nell'alveo di punibilità del reato del comma 2 anche altri soggetti.

A una prima lettura della disposizione, condotta nella prospettiva di analisi dei rapporti tra fattispecie, il reato del comma 2 sembrerebbe instaurare una peculiare relazione di "doppia" sussidiarietà, scolpita da apposite clausole espresse inserite dal legislatore delegato, che lo configurano come fattispecie sussidiaria – in quanto tale destinata a "cedere il passo", nell'ambito di un fenomeno di concorso apparente di norme – da un lato, rispetto al delitto di omissione del comma 1 (*«al di fuori dei casi indicati al comma 1»*) e, dall'altro, con altre più gravi fattispecie criminose che possano ricorrere nel caso di specie (*«Salvo che il fatto costituisca più grave reato»*).

In tale chiave di analisi, il riparto applicativo tra le varie ipotesi criminose considerate dal comma 2 – la fattispecie di "alterazione" ivi prevista, la fattispecie di "omissione" del comma 1 e gli altri più gravi reati – richiede una riflessione approfondita.

Occorre indagare, anzitutto, il rapporto con la fattispecie omissiva del comma 1, attraverso l'interpretazione del possibile significato della clausola (apparentemente) di sussidiarietà espressa contenuta nel comma 2, che identifica come principale e, dunque, prevalente nell'ambito del concorso tra norme, l'ipotesi del comma 1.

È noto, infatti, che il principio di sussidiarietà è funzionale a regolare ipotesi di concorso apparente di norme, ossia situazioni in cui un unico fatto concreto è astrattamente suscumbibile in una pluralità di fattispecie incriminatrici, di cui una soltanto, però, risulta applicabile, in ragione del fatto che punisce una forma di offesa più grave al medesimo bene giuridico ovvero l'offesa ad altri beni giuridici, ulteriori rispetto a quelli presidiati dalla fattispecie sussidiaria.

Occorre chiedersi, pertanto – sempre sul presupposto che funzione della clausola sia regolare un concorso apparente tra norme attraverso le logiche della sussidiarietà – a quale scenario applicativo faccia riferimento il lemma *«al di fuori dei casi indicati al comma 1»*, che regola il concorso con la fattispecie di "alterazione" del comma 2, in particolare interrogandosi

se/come uno stesso fatto possa al contempo presentare gli elementi costitutivi delle due ipotesi criminose.

Da questo punto di vista, sarebbe necessario, cioè, che l'omissione (come visto, totale o parziale) di «*misure tecniche di sicurezza [...] idonee a prevenire malfunzionamenti o alterazioni del funzionamento dei sistemi*» ovvero l'omissione di «*misure di sorveglianza umana*» – che determinino le situazioni di pericolo concreto descritte dal comma 1 – coesistano, fenomenologicamente, con un fatto di "alterazione" di un sistema di intelligenza artificiale ad alto rischio commesso con la medesima condotta.

L'unico scenario logicamente ipotizzabile, al fine di dare un senso – ove lo si intenda attribuire – alla clausola di sussidiarietà in esame, sarebbe quello in cui un soggetto (ad esempio, un *deployer* del sistema medico ospedaliero, che utilizza un sistema di intelligenza artificiale ad alto rischio ai fini della selezione e della prioritizzazione di pazienti destinatari di interventi di assistenza sanitaria di emergenza) abbia omissso le misure tecniche "di sicurezza" nei termini definiti dal comma 1 (ovvero abbia adottato misure che, tuttavia, non risultino adeguate) e, attraverso tale omissione, abbia determinato – dunque, "dall'interno" e non, come intende il legislatore delegato, con una «*manipolazione attiva, esterna, del sistema*» – un'alterazione del sistema stesso (che, stando all'esempio, in ragione di una "allucinazione", conduce a categorizzare come più urgente l'intervento in favore di un paziente con una patologia meno grave, ma a uno stadio più avanzato, in luogo di un altro paziente con una patologia più grave e aggressiva, tuttavia in stadio meno avanzato).

Tale ipotesi ricostruttiva, naturalmente, sarebbe plausibile a condizione di attribuire alla nozione di "alterazione" una portata ampia e onnicomprensiva, non ristretta a situazioni di manomissioni "dall'esterno" del sistema di intelligenza artificiale, in quanto tale astrattamente riconducibile anche alla condotta del soggetto "proprio" che opera sul/con il sistema di intelligenza artificiale; benché – si rammenta – nel comma 1, in effetti, il legislatore delegato abbia avuto cura di differenziare linguisticamente i «*malfunzionamenti*» dalle «*alterazioni*».

Inoltre, il presupposto implicito di simile tesi interpretativa è che l'omissione (integrante il delitto del comma 1) sia psicologicamente imputabile, nel caso concreto, a (solo) titolo di dolo, non potendo evidentemente derivare (coesistendo con essa) da un'eventuale omissione colposa – ancorché potenzialmente punibile ai sensi del comma 3 – una volontaria alterazione di un sistema di intelligenza artificiale (integrante il delitto del comma 2): il concorso apparente di norme, in questo senso, esige unitarietà dell'elemento soggettivo ascrivibile ai fatti concorrenti e, in definitiva, opererebbe soltanto tra le fattispecie dolose previste dai commi 1 e 2.

Date simili premesse, la clausola di sussidiarietà del comma 2 – ammesso e non concesso che tale sia – determinerebbe che il soggetto agente, a cui è

riferibile l'omissione delle misure "tecniche" di sicurezza e di sorveglianza umana relative a un sistema di intelligenza artificiale ad alto rischio, non possa essere punito anche per la condotta di "alterazione" del medesimo sistema.

Tuttavia, la natura principale (e, dunque, prevalente) della fattispecie del comma 1 rispetto a quella sussidiaria del comma 2 sembra essere posta in dubbio, anzitutto, dalla caratterizzazione sanzionatoria delle rispettive disposizioni incriminatrici, giacché, a ben vedere, è il delitto del comma 2 che rappresenta il reato più grave (essendo punito, nell'ipotesi base, con la reclusione da due a sei anni).

Sviluppando l'analisi in senso ulteriormente critico, poi, la tesi del rapporto di sussidiarietà tra l'ipotesi del comma 2 e quella del comma 1 perde di persuasività, dal momento che il principio di sussidiarietà, di regola, impone la sussistenza di un rapporto di progressione disvaloriale o dell'offesa (dal pericolo al danno) tra fattispecie sussidiaria e fattispecie principale, nel caso in esame non agevolmente percepibile.

Da questo punto di vista, si osserva che l'incriminazione del comma 1 risulta ancorata a una condotta essenzialmente omissiva, incardinata sul mancato compimento dell'azione giuridicamente doverosa consistente nell'adozione di misure tecniche di prevenzione di malfunzionamenti o alterazioni o di misure di sorveglianza umana – in base alle disposizioni della normativa regolatoria dedicate ai sistemi di intelligenza artificiale ad alto rischio – ossia a casistiche che (ancorché connotate da un contegno doloso, se del caso nella forma del dolo eventuale) parrebbero riferibili a carenze in fase di progettazione o, comunque, di implementazione dei sistemi di intelligenza artificiale ad alto rischio, condensabili, nella sostanza, in un rimprovero per (pur deliberata) assenza o *deficit* di "controllo preventivo".

L'ipotesi del comma 2, per converso, intende colpire condotte di dolosa (mai colposa, stante la riferibilità del comma 3 alla sola fattispecie del comma 1) alterazione dei sistemi di intelligenza artificiale ad alto rischio, come tali correlate a una diretta volontà di interferire, modificandolo anche temporaneamente, con il funzionamento di tali sistemi, le quali mostrano contiguità con altre fattispecie codicistiche (quali, ad esempio, i reati informatici di danneggiamento) e che, dal punto di vista fenomenico, sembrerebbero riconducibili piuttosto a fatti di "sabotaggio" di sistemi di intelligenza artificiale ad alto rischio.

Così inquadrata la questione circa la natura e il significato della clausola inserita nell'articolo 437-*bis*, comma 2 c.p., quest'ultima, ove analizzata nell'accennata prospettiva della sussidiarietà, sembra risentire di eccessive criticità ricostruttive, a tacer del fatto, tra l'altro, che così come congegnata pare destinata a uno spazio applicativo non particolarmente significativo.

Nell'ottica di restituire una portata effettiva al lemma normativo in uso nel comma 2, potrebbe opporsi una diversa lettura, la quale tuttavia finisce per

rinnegare la (presunta) intenzione del legislatore di disciplinare all'interno della disposizione un concorso apparente di norme.

In quest'ultima chiave di lettura, sorge il dubbio che il legislatore delegato – utilizzando un'espressione invero non appropriata – forse abbia inteso "ribadire l'ovvio", affermando che, in caso di commissione di una dolosa "alterazione", penalmente rilevante ai sensi del comma 2, da parte di un soggetto (terzo) che interviene (dall'esterno) sul sistema di intelligenza artificiale ad alto rischio, il soggetto dotato della qualifica propria (ad esempio, il provider o il *deployer*) che, in violazione del comma 1, non abbia messo in campo adeguate misure tecniche "di sicurezza" non risponde, per ciò solo, anche a titolo di concorso nella fattispecie di "alterazione" del comma 2, per averne agevolato la realizzazione.

Simile risultato interpretativo pare – come detto – al limite dell'ovvio, non foss'altro che, per imputare al soggetto "proprio" una responsabilità a titolo concorsuale con il "terzo" estraneo che opera l'alterazione, sarebbe necessario rinvenire in capo allo stesso la consapevolezza del fatto criminoso altrui e dell'apporto di un contributo causale all'alterazione del sistema, secondo le già ricordate dinamiche del dolo di concorso. Il che implica, peraltro, che potrebbero assumere rilevanza soltanto eventuali condotte omissive su base volontaria e non mere "disattenzioni" rimproverabili a titolo di colpa (ai sensi del comma 3), essendo sufficientemente consolidato l'orientamento che esclude il concorso colposo nel reato doloso.

La descritta situazione – in cui non vi sarebbe dubbio che l'*intraneus* debba rispondere anche del delitto del comma 2 realizzato dall'*extraneus* – potrebbe verificarsi, ad esempio, qualora fosse intercorso un accordo criminoso tra un *insider* (che opera per conto dell'organizzazione di un *provider* o di un *deployer*) e un altro soggetto estraneo all'organizzazione (ad esempio, un conoscente, esperto di tecniche di *cyber attack*), affinché il primo ometta una misura di sicurezza prevista a presidio del sistema di intelligenza artificiale ad alto rischio (quale lo *human oversight*), consentendo al secondo di intervenire abusivamente sul sistema medesimo, alterandolo.

Vero è che, tuttavia, se questa fosse la reale *intentio legis*, da un lato, la clausola in esame avrebbe dovuto assumere diversa connotazione linguistica (del tipo, «fuori dei casi di concorso») e, dall'altro, soprattutto, avrebbe dovuto essere inserita nel comma 1, per rinvio alla previsione del comma 2 («fuori dei casi di concorso nel reato del comma 2»).

Data l'estrema difficoltà di addivenire a un'interpretazione realmente convincente, è forse ragionevole ritenere che il legislatore delegato – lo si ripete, con linguaggio non del tutto adeguato e coerente e con strumento penalistico non appropriato – abbia semplicisticamente voluto tracciare un'*actio finium regundorum* tra le fattispecie in discorso, rimarcando che il delitto di "alterazione" del comma 2 costituisce esclusivo "fatto del terzo" (*extraneus*), nettamente distinto dall'addebito per il delitto di "omissione" del

comma 1 – applicandosi «*al di fuori dei casi*» ivi descritti – riservato al soggetto proprio della catena di valore dell'intelligenza artificiale (*intranseus*). La seconda clausola – indubbiamente di sussidiarietà espressa – aperta all'eventualità di una contestuale integrazione, insieme all'alterazione, di un «*più grave reato*», schiude alla possibilità di un concorso apparente di norme con altre fattispecie incriminatrici previste dal Codice penale – si pensi, tra gli altri, alla categoria dei reati informatici in senso stretto, intesi come *computer crimes*, quali l'accesso abusivo a un sistema informatico o telematico (articolo 615-ter c.p.) o i danneggiamenti informatici (articoli 635-bis ss. c.p.) – o da leggi speciali, che eventualmente concorrano con il fatto di alterazione del comma 2.

Il delitto del comma 2, dal punto di vista strutturale, ripropone la medesima scansione del requisito di pericolo concreto già contemplato nel comma 1 – riferito, rispettivamente, ai beni della vita o dell'incolumità (nel comma 2, soltanto) individuale (primo periodo) e dell'incolumità pubblica o della sicurezza dello Stato (secondo periodo) – che innesca un *climax* sanzionatorio ascendente (sia esterno, rispetto alle ipotesi del comma 1; sia interno, tra le ipotesi dei due periodi del comma 2) articolato nella comminatoria, nel primo caso, della reclusione da due a sei anni e, nel secondo caso, della reclusione da tre a dieci anni.

Le cornici edittali delle pene detentive ivi previste, se comparate con quelle del comma 1, forniscono ulteriore conferma circa la maggiore gravità – e la sostanziale disomogeneità tipologica – delle fattispecie di "alterazione" del comma 2.

In proposito, tra l'altro, si nota che la condotta di "alterazione" non risulta punibile, a differenza dell'ipotesi del comma 1, anche a titolo di colpa (grave, ai sensi del comma 3), a riprova che il legislatore delegato ha inteso colpire uno specifico comportamento del soggetto agente, concretantesi nella cosciente e volontaria "manomissione" del sistema di intelligenza artificiale ad alto rischio, con esclusione di fatti riconducibili a meri errori progettuali, tecnici o esecutivi in fase di intervento sui suddetti sistemi di intelligenza artificiale (ad esempio, un'attività di *fine tuning* compiuta maldestramente dal *deployer* o un *retraining* poco "oculato", effettuato attraverso l'utilizzo di *dataset* non accurati o aggiornati).

4.3. Responsabilità delle persone fisiche e *ne bis in idem* europeo.

Un ultimo profilo che, per completezza, merita di essere approfondito concerne la potenziale interferenza delle nuove incriminazioni introdotte dal d.lgs. 160/2026 con il divieto di *bis in idem*, nella sua accezione sostanziale di diritto europeo e convenzionale, in considerazione del fatto che i suddetti delitti sono destinati a inserirsi in un contesto sanzionatorio già assai "affollato", in particolare di illeciti amministrativi "propri" posti a presidio dell'osservanza della normativa regolatoria dell'AI Act.

Potrebbe darsi l'eventualità, cioè, che il giudice penale si trovi a dover tenere in considerazione, secondo le ormai esplorate logiche del *ne bis in idem* (articolo 4 Prot. 7 CEDU e articolo 50 CDFUE), la precedente irrogazione di una sanzione amministrativa – applicata nei confronti del/i trasgressore/i persona fisica/che dalle competenti autorità di settore per gli illeciti riferibili alla violazione di una o più disposizioni regolatorie dell'AI Act – in fase di determinazione della pena per uno dei delitti dell'articolo 437-*bis* c.p., rispetto ai quali abbia riconosciuto la responsabilità penale della/e medesima/e persona/e fisica/che.

Occorre preliminarmente delimitare, tuttavia, l'ipotetico campo di applicazione del divieto in analisi rispetto alle fattispecie criminose dell'articolo 437-*bis* c.p.

In proposito, si rammenta sinteticamente che – secondo l'ormai consolidata giurisprudenza della Corte EDU e della Corte di giustizia dell'Unione europea, recepita anche dalle corti nazionali – il divieto di *bis in idem* consacrato nell'art. 4 Prot. 7 CEDU e nell'articolo 50 CDFUE trova applicazione in presenza di un "doppio binario sanzionatorio" previsto nell'ordinamento nazionale, basato sulla coesistenza, in relazione allo stesso fatto (*idem factum*), di illeciti (e correlative sanzioni) di carattere formalmente "amministrativo" che siano qualificabili come "sostanzialmente penali" con altri illeciti (e sanzioni) di carattere penale in senso proprio.

Si ricorda, sul punto, che la medesimezza del fatto, quale presupposto logicamente prioritario di operatività del divieto – secondo granitica giurisprudenza europea (tra le altre, la sentenza della Corte EDU *Zolotukhin c. Russia*) – non deve essere rapportata all'astratta qualificazione giuridica degli illeciti (*idem legale*), bensì alla materiale coincidenza del "fatto concreto" (*idem factum*) oggetto di contestazione, inteso come "nucleo storico" della condotta e del contesto spazio-temporale di sua verifica. Il secondo presupposto applicativo del divieto attiene alla qualificazione giuridica dell'illecito amministrativo (e delle relative sanzioni) come "sostanzialmente penale", alla luce dei criteri enunciati dalla celebre sentenza *Engel* della Corte EDU e successivamente sviluppati, non senza incertezze, dalla giurisprudenza successiva anche della Corte di giustizia dell'Unione europea, in funzione di smascheramento di "frodi delle etichette" atte a celare, sotto un *nomen iuris* diverso, illeciti che (per ambito di applicazione generale, interessi tutelati, finalità, afflittività in concreto e procedure applicative delle connesse sanzioni) assumano carattere "punitivo".

Occorre, altresì – sulla scorta dell'insegnamento della Corte di giustizia nelle cause riunite *Orsi* e *Baldetti* – un rapporto di identità soggettiva tra il/i soggetto/i destinatario/i delle rispettive sanzioni irrogate all'esito dei due diversi procedimenti, amministrativo e penale.

La valutazione in ordine a una potenziale duplicazione sanzionatoria per lo stesso fatto, censurabile ai sensi delle richiamate disposizioni della CEDU e

della CDFUE, infine, è stata ancorata dalla giurisprudenza europea, parimenti recepita da quella nazionale, al *test* di “sufficiente connessione sostanziale e temporale” (*sufficiently close connection in substance and time*, affermata dall’altrettanto nota sentenza *A e B c. Norvegia*) tra i due procedimenti che hanno condotto all’irrogazione delle sanzioni integranti il “binario”, condizione che, ove sussistente, rende legittimo il regime nazionale di “doppio binario” inteso alla stregua di un sistema sanzionatorio “integrato”. Sotto il profilo della qualificazione giuridica, non pare esservi dubbio che le sanzioni amministrative pecuniarie (ma non solo) derivanti dal combinato disposto dell’articolo 99 AI Act e dell’articolo 23 d.lgs. 179/2026 possano essere qualificate come “sostanzialmente penali” – in particolare in relazione al loro carattere punitivo – sol considerando la natura “draconiana” dei massimi edittali definiti su base numerica e, in ogni caso, del metodo di calcolo della sanzione parametrato al fatturato totale mondiale annuo dell’impresa, che presuppone quale risultante un importo pecuniario addirittura superiore.

Quanto alla ricorrenza del requisito di “identità” del fatto sanzionato in via amministrativa e penale, si osserva che, a ben vedere, l’eventualità per il giudice penale di dover procedere con lo “scioglimento” del cumulo sanzionatorio sembra doversi porre, non per tutte le ipotesi criminose dell’articolo 437-*bis* c.p., ma soltanto in relazione ai delitti di omessa adozione di misure tecniche di sicurezza e di sorveglianza umana, di cui all’articolo 437-*bis*, commi 1 e 3 c.p.

Esse, invero, rappresentano le uniche due fattispecie che – ove comparate alla materialità degli illeciti amministrativi previsti a presidio della normativa dell’AI Act – paiono integrare il presupposto dell’*idem factum*, poiché, se gli illeciti amministrativi in materia di intelligenza artificiale hanno chiara natura “ingiunzionale”, essendo preposti a colpire lo specifico profilo di violazione delle disposizioni regolatorie dell’AI Act, i delitti di “omissione” delle misure di sicurezza e di sorveglianza umana, nel caso concreto, potrebbero finire per punire proprio il medesimo profilo di violazione, in relazione alla mancata adozione (omissione) delle “misure” cogenti previste dall’AI Act, oggetto in prima battuta di tutela amministrativa.

Naturalmente, si potrebbe obiettare che l’identità del fatto non è piena, giacché i due delitti prevedono il requisito di pericolo concreto – sia esso ritenuto tale (invece che condizione obiettiva di punibilità) – quale ulteriore elemento costitutivo, che differenzerebbe l’illecito penale da quello amministrativo.

Ricorrerebbe, nel caso di specie, anche l’ulteriore requisito di coincidenza soggettiva tra i soggetti sanzionati, dal momento che vi sarebbe, da un lato, il/i trasgressore/i cui l’illecito è contestato ai sensi dell’articolo 14 l. 689/1981 (oltre che, in un contesto aziendale – come si approfondirà nel seguito – in cui «la violazione è commessa dal rappresentante o dal dipendente», la

responsabilità solidale, ai sensi dell'articolo 6, comma 3 l. 689/1981, di «*una persona giuridica o di un ente privo di personalità giuridica*»), e, dall'altro, l'imputato/i persona/e fisica/che – che verosimilmente, per il tipico effetto di “risalita verso l'alto” delle responsabilità, saranno i soggetti apicali dell'ente – nei cui confronti è stata esercitata l'azione penale per i delitti di “omissione”. Il rischio di una “sovrapposizione” sanzionatoria tra il comparto dell'Al Act e i delitti codicistici di “omissione” – come visto – appare concreto e chiamerà certamente in futuro i giudici di merito e di legittimità quanto meno a porsi il problema del “doppio binario”.

Per converso, il delitto di “alterazione” dell'articolo 437-*bis*, comma 2 c.p. – e, parimenti, il delitto dell'articolo 612-*quater* c.p., cui oltre si accennerà – sembrano esulare dal campo applicativo della garanzia eurounitaria e convenzionale, in ragione della propria evidente alterità strutturale e fenomenica rispetto agli illeciti amministrativi di settore.

AmMESSO e non concesso che ricorra una situazione riconducibile al divieto di *bis in idem*, a complicare ulteriormente il nodo applicativo, occorre tenere presente che, nel decidere la pena da applicare per i delitti di “omissione” dell'articolo 437-*bis* c.p., il giudice penale sarà chiamato a operare uno “scomputo” di carattere eterogeneo, tra sanzioni amministrative pecuniarie e pena della reclusione (che, se del caso, potrebbe essere astrattamente sostituita, entro il limite di un anno, con la pena pecuniaria sostitutiva ai sensi dell'articolo 53 l. 689/1981), nella totale assenza sia di una norma che legittimi espressamente il giudicante a realizzare (*an*) forme di “compensazione” sanzionatoria sia di criteri che regolino le modalità (*quomodo*) con cui procedere al calcolo della sanzione finale in ambito penale.

Sotto questo profilo, infatti, il legislatore delegato – non è dato sapere se consapevolmente o meno – ha evitato di inserire nella trama normativa dei decreti legislativi attuativi della l. 132/2025 una disposizione sulla falsariga di quelle già previste, ad esempio, nel settore finanziario (articolo 187-*terdecies* d.lgs. 58/1998, come sostituito dal d.lgs. 107/2018 di adeguamento al regolamento (UE) n. 596/2014) o, più di recente, in ambito fiscale (articolo 21-*ter* d.lgs. 74/2000, introdotto dal d.lgs. 87/2024 di revisione del sistema sanzionatorio tributario, “sopravvissuto” all'intervento sostitutivo del d.lgs. 173/2024) che consentano all'autorità procedente nella rispettiva sede (amministrativa o penale) di tenere in considerazione la precedente irrogazione, per lo stesso fatto, a seconda dei casi, di una sanzione di carattere pecuniario, detentivo o interdittivo, da scomputare – nel procedimento che per secondo interviene a irrogare la sanzione – da quella destinata a essere applicata nella sede susseguente.

Sarebbe stata opportuna e doverosa, in tal senso – ponendo mente alle previsioni dell'Al Act (si vedano, in particolare, l'articolo 99, par. 1, secondo periodo Al Act, che fa riferimento a sanzioni «*effettive, proporzionate e*

dissuasive», nonché, ancor più specificamente, il considerando n. 168, che chiama gli Stati membri ad adottare, per garantire l'*enforcement* del regolamento, «*sanzioni effettive, proporzionate e dissuasive in caso di violazione, anche nel rispetto del principio ne bis in idem*») – l'introduzione di una disposizione modellata sugli archetipi dei menzionati articoli 187-terdecies d.lgs. 58/1998 e 21-ter d.lgs. 74/2000, in forza della quale, quando «*per lo stesso fatto*» è stata applicata, a carico del reo, dell'autore della violazione o dell'ente (o meglio, come previsto nell'articolo 21-ter d.lgs. 74/2000, del «*soggetto*») una sanzione amministrativa pecuniaria ovvero una sanzione penale o una sanzione amministrativa dipendente da reato, l'autorità giudiziaria o le autorità competenti in base alle disposizioni nazionali di adeguamento dell'AI Act «*tengono conto, al momento dell'irrogazione delle sanzioni di propria competenza, delle misure punitive già irrogate*», «*con provvedimento o con sentenza assunti in via definitiva*», «*al fine di ridurre la relativa misura*».

Allo stato, peraltro – in assenza del supporto del dato normativo – non mancano soluzioni giurisprudenziali mutuabili, elaborate nel corso degli anni di applicazione del divieto di *bis in idem* in relazione ad altri plessi normativi regolatori (partendo proprio da quello finanziario e fiscale, sui cui hanno fatto diretta presa gli insegnamenti della pronuncia "capostipite" della Corte EDU *Grande Stevens c. Italia*), che hanno consentito al giudice penale, in via interpretativa, ad esempio facendo perno sui criteri di commisurazione della pena dettati dall'articolo 133 c.p., di operare, se del caso previa conversione della sanzione amministrativa "sostanzialmente penale" irrogata dall'autorità regolatoria, una riduzione, se non un totale azzeramento, della pena principale da irrogare all'esito del procedimento penale.

Quale che sia la soluzione applicativa che si intenderà seguire, è evidente che – in un settore come quello dell'intelligenza artificiale, caratterizzato da un elevato grado di incertezza tecnica e scientifica e presidiato da sanzioni amministrative "proprie" che possono raggiungere livelli di afflittività molto elevati – è forte l'esigenza di rinvenire nelle pieghe dell'ordinamento uno strumento atto a bilanciare, riportando a equità e a proporzionalità, le sanzioni che siano irrogate nell'ambito del complessivo binario sanzionatorio "integrato" contemplato dall'ordinamento nazionale.

4.4. Annotazioni conclusive sul delitto dell'articolo 437-bis c.p.

La panoramica svolta sul nuovo delitto codicistico e sulle numerose criticità ermeneutiche che esso pone sul tavolo mette in luce, conclusivamente, che il legislatore delegato ha adottato un approccio, al contempo, "minimale" ed estensivo nell'attuazione dei principi e dei criteri direttivi posti dalla l. 132/2025.

Seppur siano state previste nell'articolo 437-bis c.p. «*autonome fattispecie di reato, punite a titolo di dolo o di colpa*», il fuoco dell'incriminazione non

sembra del tutto in linea con le indicazioni della l. 132/2025, che richiede, a ben vedere, la punizione di condotte incentrate *«sull'omessa adozione o sull'omesso adeguamento di misure di sicurezza per la produzione, la messa in circolazione e l'utilizzo professionale di sistemi di intelligenza artificiale, quando da tali omissioni deriva pericolo concreto per la vita o l'incolumità pubblica o individuale o per la sicurezza dello Stato»* (articolo 24, comma 5, lett. b)).

Ferme le considerazioni svolte in relazione alla soltanto apparente limitazione dell'oggetto materiale delle condotte (riferite ai soli sistemi di intelligenza artificiale ad alto rischio) – cui si rinvia – si osserva come nell'articolo 437-bis c.p. non figurino fattispecie di reato incardinate *«sull'omesso adeguamento»* delle misure di sicurezza dei sistemi di intelligenza artificiale ad alto rischio, salvo ritenere che l'incriminazione del comma 1, in ragione dell'implicito richiamo della normativa dell'AI Act in tema di misure (anche) correttive, finisca per assorbire, coprendo, il disvalore del fatto di "mancato adeguamento" richiesto dalla l. 132/2025.

Dall'altra parte, non sembra trovarsi all'interno della l. 132/2025 alcun riferimento espresso all'introduzione, tra i fatti sanzionabili penalmente, alla condotta di "alterazione", che invece compare nell'articolo 437-bis, comma 2, profilandosi sotto questo aspetto un possibile rischio di eccesso di delega censurabile in sede di giudizio di costituzionalità (per violazione degli articoli 76 e 25 Cost.).

Al contempo, in una prospettiva più ampia, come si è già anticipato, il legislatore delegato non si è "spinto" fino all'obiettivo – pur prefigurato dalla l. 132/2025 – di precisare autonomi e speciali *«criteri di imputazione della responsabilità penale delle persone fisiche [...] per gli illeciti inerenti a sistemi di intelligenza artificiale»*, i quali, tra l'altro, avrebbero dovuto tenere in considerazione, in punto di colpevolezza e, in particolare, di "dominabilità" dei fattori di rischio, il *«livello effettivo di controllo dei sistemi predetti da parte dell'agente»* (articolo 24, comma 5, lett. c)).

La definizione dei criteri di imputazione della responsabilità penale connessa ai sistemi di intelligenza artificiale, come ricordato, costituisce oggi più che mai un "cantier aperto", nell'ambito del quale – tra soluzioni che spaziano dalla negazione di autonomi margini applicativi del diritto penale (in favore di schemi oggettivistici di allocazione delle responsabilità) fino all'estremo del riconoscimento di una responsabilità propria della macchina (*machina delinquere potest*) – la copiosa letteratura specialistica ha già messo in guardia dalle criticità connesse alla costruzione di un puntuale "statuto penale" dell'intelligenza artificiale, che risulti rispettoso dei canoni costituzionali in materia penale (primo fra tutti il principio di personalità della responsabilità penale, intesa come responsabilità per fatto proprio e colpevole, di cui all'articolo 27, comma 1 Cost.).

Un simile innesto normativo – che condurrebbe a una vera e propria rivoluzione “copernicana” del diritto penale e delle sue categorie dogmatiche tradizionali – richiede evidentemente maggiori tempi di maturazione e una più meditata riflessione sulle direttrici evolutive degli attuali paradigmi sanzionatori rispetto a fatti che implicano, nella propria realizzazione, il coinvolgimento di sistemi di intelligenza artificiale.

5. Rischi “da intelligenza artificiale” e responsabilità “da reato” dell’ente (d.lgs. 231/2001).

Sul versante della responsabilità amministrativa dipendente da reato degli enti (d.lgs. 231/2001) è possibile apprezzare un duplice ordine di modifiche – dirette e indirette (o riflesse) – che scaturiscono dal combinato delle disposizioni della l. 132/2025 e del d.lgs. 160/2026, destinate a impattare profondamente sull’approccio alla *compliance* preventiva delle imprese rispetto ai rischi (penali e, ancor prima, regolatori) connessi ai sistemi di intelligenza artificiale.

5.1. Il “catalogo” dei reati presupposto: le modifiche dirette.

La prima modifica diretta (articolo 15) consiste nell’estensione del “catalogo” dei reati presupposto della responsabilità dell’ente, attraverso l’inserimento di un inedito articolo 25-*vicies* d.lgs. 231/2001, dedicato ai «*Reati commessi con l’uso di sistemi di intelligenza artificiale*».

La disposizione introduce come reato presupposto della responsabilità dell’ente, in primo luogo, le fattispecie (sebbene l’articolo 25-*vicies* d.lgs. 231/2001 si riferisca a «*il delitto*») previste dall’articolo 437-*bis*, commi 1, 2 e 3 c.p., le quali, pertanto, ove commesse dai soggetti qualificati previsti dall’articolo 5 d.lgs. 231/2001 (“apicali” e/o “sottoposti”), nell’interesse o a vantaggio dell’ente, in presenza di profili di “colpa di organizzazione”, potranno determinare un’inculpazione anche a carico del soggetto metaindividuale.

Dal punto di vista diacronico, tuttavia, si ribadisce quanto sopra precisato circa le prospettive di immediata contestazione delle fattispecie di reato presupposto dell’articolo 437-*bis*, commi 1 e 3 c.p. (omessa adozione delle misure tecniche di sicurezza e delle misure di sorveglianza umana, dolosa e colposa) e, con esse, dei connessi illeciti amministrativi dell’ente, destinati a seguire il regime “prorogato” di applicabilità dei requisiti normativi dei sistemi di intelligenza artificiale ad alto rischio (*retro*, paragrafo 2).

A cascata, infatti, in forza del principio di (doppia) legalità posto dall’articolo 2 d.lgs. 231/2001, sarà possibile avanzare un addebito nei confronti dell’ente ai sensi del nuovo articolo 25-*vicies* d.lgs. 231/2001 – eccezion fatta per l’ipotesi di “alterazione” prevista dall’articolo 437-*bis*, comma 2 c.p., che non risente del *timing* normativo e che, dunque, sarà immediatamente

contestabile – soltanto a partire, rispettivamente, dal 2 dicembre 2027 e dal 2 agosto 2028, a seconda del sistema di intelligenza artificiale ad alto rischio interessato dalla condotta omissiva.

Ad ogni modo, la novella normativa – allorquando a pieno regime – segnerà un punto di svolta per la *compliance* preventiva delle imprese ricadenti nel campo di applicazione definito dall'articolo 1 d.lgs. 231/2001 e, in particolare, quelle che operano lungo la "filiera" dell'intelligenza artificiale (soprattutto, in qualità di *provider* e di *deployer*), le quali potranno essere coinvolte nel procedimento di accertamento della relativa responsabilità "da reato" non solo – si badi bene – per condotte criminali "devianti" di natura dolosa, ma anche per fatti meramente colposi (in particolare, omissioni rimproverabili a titolo di colpa grave) commessi nel fisiologico esercizio delle attività aziendali correlate all'intero "ciclo di vita" dei sistemi di intelligenza artificiale ad alto rischio.

Rispetto all'illecito dipendente dal reato (*rectius*, dai reati) dell'articolo 437-bis c.p., il nuovo articolo 25-*vicies* d.lgs. 231/2001 prevede un trattamento sanzionatorio che, dal punto di vista della sanzione pecuniaria, spazia da un minimo di seicento a un massimo di mille quote, nonché (comma 3) l'applicabilità delle sanzioni interdittive previste dall'articolo 9, comma 2, lett. b), c), d) ed e) d.lgs. 231/2001.

In prima battuta, si osserva che il legislatore delegato ha optato per un approccio assai afflittivo verso l'ente che abbia consentito, per propria "colpa in organizzazione", la commissione dei delitti dell'articolo 437-bis c.p., dal momento che il massimo edittale per l'illecito amministrativo dipendente da reato (mille quote), da un lato, rappresenta il numero massimo delle quote applicabile *ex lege* (articolo 10, comma 2 d.lgs. 231/2001) e, dall'altro, se non proprio un *unicum* nel quadro regolatorio del d.lgs. 231/2001, costituisce senza dubbio un caso di particolare pregnanza sanzionatoria, che si pone appena un gradino al di sotto della previsione draconiana della sanzione "fissa" di mille quote prevista dall'articolo 25-*septies*, comma 1 d.lgs. 231/2001 in caso di omicidio colposo (commesso con violazione delle norme sulla tutela della salute e sicurezza sul lavoro) aggravato dalla violazione della "speciale" disposizione del d.lgs. 81/2008.

In secondo luogo, si nota che l'articolo 25-*vicies* d.lgs. 231/2001 dispone un trattamento sanzionatorio pecuniario unitario per tutte e tre le diverse fattispecie incriminatrici dell'articolo 437-bis c.p. costituenti reato presupposto – l'omissione di misure di "sicurezza", dolosa e colposa (commi 1 e 3) e l'alterazione (comma 2) – laddove sarebbe stato lecito attendersi un trattamento differenziato, in rapporto di scalarità continua con il disvalore penale associato alle diverse ipotesi delittuose.

In tale direzione, un utile paradigma di riferimento ai soli fini sanzionatori – nella diversità strutturale delle fattispecie sottostanti – sarebbe stato rinvenibile, ad esempio, nell'articolo 25-*undecies* d.lgs. 231/2001, in materia

di reati ambientali, che rappresentano l'unica sede del d.lgs. 231/2001 in cui sono previste e sanzionate fattispecie di natura delittuosa, punibili sia a titolo di dolo che di colpa.

In particolare, il comma 1 della disposizione, nelle lett. *a)*, *b)* contempla, rispettivamente, i delitti di inquinamento ambientale e di disastro ambientale dolosi (punibili ai sensi degli articoli 452-*bis* e 452-*quater* c.p.) e nella lett. *c)* le corrispondenti fattispecie colpose (punibili ai sensi dell'articolo 452-*quinqüies* c.p.), assoggettando – in rapporto di accennata scalarità – queste ultime (meno gravi) a una sanzione pecuniaria da duecento a cinquecento quote, la prima (dal disvalore intermedio), a una sanzione pecuniaria da quattrocento a seicento quote e la seconda (dal disvalore massimo) a una sanzione pecuniaria da seicento a milleduecento quote (valore in origine pari a novecento quote e solo di recente modificato, in incremento e addirittura oltre il numero massimo di legge delle quote, in virtù del d.l. 116/2025, conv. con mod. in l. 147/2025, c.d. decreto Terra dei fuochi).

La stessa scalarità avrebbe potuto trovare applicazione prevedendo una sanzione pecuniaria, secondo cornici edittali in ordine crescente per le varie ipotesi dell'articolo 437-*bis* c.p., per l'omissione colposa del comma 3 c.p., per la corrispondente ipotesi dolosa del comma 1 e, infine, per la fattispecie di "alterazione" del comma 2. Al più, il legislatore delegato avrebbe potuto differenziare il trattamento sanzionatorio pecuniario per i delitti previsti dal comma 1 e dal comma 2, prevedendo una riduzione frazionaria (in linea con quanto previsto per la pena delle persone fisiche) per il delitto del comma 3. È evidente che una comminatoria *flat*, come quella contenuta nell'articolo 25-*vicies* d.lgs. 231/2001, rischia di esporre la disposizione a potenziali censure di legittimità costituzionale in relazione, tra gli altri, al parametro di uguaglianza-ragionevolezza (articolo 3 Cost.), le quali, tuttavia, potrebbero essere superate in via applicativa, oltre che argomentativa, facendo perno sulla natura "flessibile" della sanzione pecuniaria in concreto irrogabile, in forza del meccanismo di commisurazione bifasico previsto dal d.lgs. 231/2001.

Così, l'apparente "schiacciamento" della sanzione pecuniaria rispetto a reati presupposto (e relativi illeciti dell'ente) recanti un disvalore di fattispecie eterogeneo potrebbe essere recuperato, in sede di esercizio del potere di dosimetria sanzionatoria del giudice penale, mediante la scelta di applicare un numero inferiore di quote e più vicino al minimo edittale (seicento) e/o un valore più basso della singola quota, prossimo al valore minimo di legge (258 Euro) per l'illecito dipendente dal reato meno grave – l'omissione colposa di cui all'articolo 437-*bis*, comma 3 c.p. – e, via via, modulando la sanzione pecuniaria con gli accennati fattori del calcolo bifasico, garantire la proporzionalità e la giustizia sostanziale della risposta sanzionatoria per gli illeciti dipendenti dalle fattispecie gradatamente più gravi – l'omissione

dolosa e l'alterazione di cui, rispettivamente, all'articolo 437-*bis*, commi 1 e 2 c.p.

Quanto all'impatto economico in concreto delle sanzioni pecuniarie, sempre e comunque applicabili in caso di riconoscimento della responsabilità (articolo 10, comma 1 d.lgs. 231/2001), astrattamente l'ente – al netto di eventuali, ma non da escludere, definizioni alternative della regiducanda penale (ad esempio, attraverso applicazione della sanzione su richiesta ai sensi dell'articolo 63 d.lgs. 231/2001) e del ricorrere di una o più circostanze attenuanti previste dall'articolo 12 d.lgs. 231/2001 – rischia l'irrogazione di una sanzione pecuniaria:

- nel minimo, compresa nel *range* tra 154.800 Euro e 925.800 Euro, derivante dalla moltiplicazione del numero minimo di quote previste per l'illecito amministrativo (seicento) per il valore minimo (258 Euro) e massimo (1549 Euro) di ciascuna quota (valore della quota che sarà definito in concreto dal giudice penale in base ai criteri di commisurazione dell'articolo 11 d.lgs. 231/2001);
- nel massimo, compresa nel *range* tra 258.000 Euro e 1.549.000 Euro, derivante dalla moltiplicazione del numero massimo di quote previste per l'illecito amministrativo (mille) per i menzionati valori, minimi e massimi, di ciascuna quota.

Si nota, per ultimo, che, in sede di commisurazione della sanzione pecuniaria a carico dell'ente – non diversamente da quanto già osservato nei rapporti tra i delitti di "omissione" dell'articolo 437-*bis* c.p. ascrivibili alla persona fisica e gli illeciti amministrativi di settore, oggi disciplinati nel d.lgs. 179/2026 (*retro*, paragrafo 4.3.) – il giudice penale potrebbe trovarsi nella "scomoda" posizione, in assenza di parametri espressi per compiere l'operazione, di valutare la doverosità, secondo le logiche del *ne bis in idem* di matrice europea e sovranazionale e in un'ottica di giustizia sostanziale, di uno "scomputo" dell'eventuale precedente sanzione amministrativa irrogata dall'autorità competente per il concorrente (con il reato presupposto e il rispettivo illecito dell'ente) profilo di violazione della disciplina regolatoria in materia di intelligenza artificiale prevista dall'AI Act (articolo 99) e dalla normativa nazionale di adeguamento (articolo 23 d.lgs. 179/2026).

Il tema di una "compensazione" giudiziale – nel caso di specie "omogenea", essendo destinata a intervenire tra sanzioni di pari natura pecuniaria – si pone, anche con riferimento alla posizione dell'ente-organizzazione, soltanto in relazione all'illecito dipendente dai delitti di omessa adozione di misure tecniche di sicurezza e di sorveglianza umana, di cui all'articolo 437-*bis*, commi 1 e 3 c.p., le quali, come detto, rappresentano le uniche due fattispecie suscettibili di integrare il presupposto dell'*idem factum*, con esclusione, per le ragioni già esposte, del contiguo delitto di "alterazione" (articolo 437-*bis*, comma 2 c.p.) e del delitto di illecita diffusione (articolo 612-*quater* c.p.).

Sotto questo profilo, a essere poste in ipotetica comparazione, ai fini dell'operazione di "scioglimento" del cumulo, sarebbero:

- da un lato, le sanzioni pecuniarie da irrogare in sede penale (obbligatoriamente, in caso di accertamento della responsabilità) a carico dell'ente ai sensi del d.lgs. 231/2001, in relazione all'illecito – dipendente dai delitti di cui all'articolo 437-*bis*, commi 1 e 3 c.p. concretizzatisi nell'omessa adozione di misure imposte da una o più disposizioni regolatorie dell'AI Act aventi rilevanza anche amministrativa – commesso da apicali e/o sottoposti dell'ente (già sanzionati o sanzionabili in via amministrativa), nell'interesse o a vantaggio dello stesso, in presenza dell'elemento costitutivo della "colpa di organizzazione";
- dall'altro, le sanzioni amministrative pecuniarie per la/e violazione/i dell'AI Act – coincidente/i (*idem factum*) con il profilo di omissione contestato all'ente ai sensi dell'articolo 25-*vicies* d.lgs. 231/2001 in relazione all'articolo 437-*bis* c.p. – irrogate non direttamente all'ente quale trasgressore, ma quale soggetto obbligato in solido ai sensi dell'articolo 6, comma 3 l. 689/1981.

Tale operazione implica, dal punto di vista giuridico, di poter qualificare come sanzioni "sostanzialmente penali":

- le sanzioni amministrative pecuniarie irrogate per violazione/i dell'AI Act dalle competenti autorità di settore anche a carico dell'ente, oltre che del/i trasgressore/i persona/e fisica/che, sul presupposto – pur non pacifico in giurisprudenza, benché in taluni casi riconosciuto – che la responsabilità solidale prevista dall'articolo 6, comma 3 l. 689/1981 abbia natura "punitiva";
- le sanzioni da applicare a carico dell'ente nel procedimento di accertamento della responsabilità "da reato", sul presupposto – che non è possibile in questa sede esaminare compiutamente – che nel modello di tutela sanzionatoria del d.lgs. 231/2001 siano rinvenibili i tratti tipici della *matière penale*.

Ipotizzata la ricorrenza anche degli altri requisiti e condizioni che fondano l'applicazione del divieto di matrice europea (tra cui la mancanza di *sufficiently close connection in substance and time* tra i procedimenti sanzionatori), il giudice penale chiamato a decidere della responsabilità dell'ente ai sensi del d.lgs. 231/2001 nel *simultaneus processus* dovrebbe porsi il problema di un'eventuale decurtazione (se non del totale assorbimento) della sanzione pecuniaria applicabile all'ente, in ragione – sia concessa l'espressione – del "pre-sofferto" dello stesso ente nella precedente sede amministrativa.

Si richiamano in questa sede le considerazioni in precedenza svolte in ordine alla mancata previsione da parte del legislatore delegato di un meccanismo di scomputo, che possa valere anche in relazione alle sanzioni nei confronti

dell'ente nella sede susseguente (a seconda della scansione temporale dei procedimenti: il procedimento di irrogazione delle sanzioni amministrative ovvero il procedimento penale di accertamento della responsabilità dell'ente ai sensi del d.lgs. 231/2001), che avrebbe potuto trarre ispirazione dal prototipo esistente in materia finanziaria e tributaria.

Sul punto, si rammenta che, in ambito finanziario, l'articolo 187-*terdecies* d.lgs. 58/1998 prende espressamente in considerazione il tema del cumulo sanzionatorio a carico della *societas*, per l'ipotesi in cui sia stata applicata «a carico [...] dell'ente [...] una sanzione amministrativa pecuniaria [...] o una sanzione amministrativa dipendente da reato», disponendo che l'autorità competente di settore (Consob) e l'autorità giudiziaria «*tengono conto, al momento dell'irrogazione delle sanzioni di propria competenza, delle misure punitive già irrogate*».

La disposizione, tra l'altro, in ragione della neutralità della terminologia utilizzata, con riferimento alle sanzioni "penali", consente astrattamente di attrarre nell'operazione di scomputo non soltanto la sanzione pecuniaria (obbligatoriamente) irrogata all'ente all'esito del procedimento di cui al d.lgs. 231/2001, ma anche l'eventuale sanzione interdittiva applicate in tale contesto, in quanto comunque lessicalmente riconducibile alle nozioni di «*sanzione amministrativa dipendente da reato*» e di «*misure punitive già irrogate*».

Ancor più ampio appare il meccanismo di scomputo previsto in ambito tributario, dove l'articolo 21-*ter* d.lgs. 74/2000, nel contemplare la posizione dell'ente, non soltanto fa riferimento, quanto alle sanzioni "penali", all'onnicomprendente nozione di «*sanzione amministrativa dipendente da reato*», ma addirittura sembra estendere le prospettive di scioglimento del cumulo, quanto alle sanzioni "amministrative", alludendo al concetto generico di «*sanzione amministrativa*», non qualificata esclusivamente come "pecuniaria".

Ebbene, ove il legislatore delegato avesse inteso replicare una disposizione di simile tenore – vuoi nel d.lgs. 160/2026 vuoi nel d.lgs. 179/2026 – rispetto alla posizione dell'ente avrebbero potuto essere valorizzate, al fine di operare una riduzione delle sanzioni da irrogare nella sede susseguente, a seconda dei rapporti cronologici tra procedimenti, le sanzioni amministrative applicate dall'autorità competente di settore per le violazioni dell'Al Act, siano esse «*pecuniarie*» (articolo 23 d.lgs. 179/2026) o anche, eventualmente, «*non pecuniarie*» (articolo 22 d.lgs. 179/2026, ponendosi semmai in tal caso il problema di come operare concretamente lo scomputo), ovvero le sanzioni pecuniarie irrogate dal giudice penale per gli illeciti di cui all'articolo 25-*vicies* d.lgs. 231/2001, nonché le eventuali sanzioni interdittive per essi applicate (rispetto alle quali si pone analoga questione di scomputabilità da sanzioni di carattere strettamente pecuniario, operazione per la quale, tuttavia, potrebbe rinvenirsi un utile modello di riferimento nell'articolo 78 d.lgs.

231/2001, il quale, per la fase di esecuzione delle sanzioni, contempla, in caso di "condotte riparatorie", un meccanismo di conversione della sanzione interdittiva in sanzione pecuniaria, che è effettuato «*determinando l'importo della sanzione pecuniaria in una somma non inferiore a quella già applicata in sentenza e non superiore al doppio della stessa*»).

Sotto quest'ultimo profilo, infatti, si osserva che per l'ente, al netto della pur preoccupante prospettiva sanzionatoria pecuniaria (anche, ma non solo all'esito del procedimento di cui al d.lgs. 231/2001), il vero "spauracchio" per l'organizzazione è rappresentato dall'applicabilità (articolo 25-*vicies*, comma 3 d.lgs. 231/2001) delle sanzioni interdittive previste dall'articolo 9, comma 2 d.lgs. 231/2001 e, nel dettaglio:

- la sospensione o la revoca delle autorizzazioni, licenze o concessioni funzionali alla commissione dell'illecito (lett. b);
- il divieto di contrattare con la pubblica amministrazione, salvo che per ottenere le prestazioni di un pubblico servizio (lett. c);
- l'esclusione da agevolazioni, finanziamenti, contributi o sussidi e l'eventuale revoca di quelli già concessi (lett. d);
- il divieto di pubblicizzare beni o servizi (lett. e).

Il legislatore delegato, piuttosto sorprendentemente – forse per non gravare oltremodo l'impresa, in particolare quella operante nel settore dello sviluppo di "soluzioni" di intelligenza artificiale o che, comunque, fonda il proprio *business* in maniera preponderante su processi *AI-oriented*, compromettendone la continuità aziendale – ha deciso di non prevedere, tra le sanzioni interdittive applicabili, l'interdizione (anche parziale) dall'esercizio dell'attività (lett. a); scelta che certamente farà discutere in chiave di finalità preventiva e deterrente circa lo strumento sanzionatorio apprestato da una normativa, come il d.lgs. 231/2001, caratterizzata nella propria storia applicativa da tutta una serie di "debolezze" (genetiche e sopravvenute) che ne hanno minato fino a oggi la "presa" in campo giudiziario.

L'effetto di particolare *metus* annesso alla comminatoria interdittiva – com'è noto – discende dal fatto che le sanzioni interdittive, a differenza di quelle pecuniarie, non colpiscono l'ente soltanto "alla fine della corsa", ossia dopo l'esaurimento della vicenda penale che lo riguarda (in tutti gli eventuali gradi di giudizio, di merito e di legittimità, destinati a protrarsi negli anni) e soltanto una volta che sarà concretamente messa in moto la macchina esecutiva degli uffici di Procura (il che, notoriamente, avviene a distanza di anni), bensì fin dal principio del procedimento penale, addirittura nelle battute iniziali delle indagini preliminari, ove le stesse possono trovare manifestazione sotto forma di misure cautelari interdittive, in costanza dei presupposti di legge previsti dall'articolo 45 d.lgs. 231/2001.

Ciò significa che l'ente, potenzialmente, potrebbe trovarsi, da un giorno all'altro, menomato nella propria operatività dall'esecuzione di un'ordinanza del giudice (per le indagini preliminari o delle fasi successive), che, in

accoglimento della richiesta del pubblico ministero, abbia disposto, pur all'esito del contraddittorio difensivo previsto dall'articolo 47 d.lgs. 231/2001, il divieto di contrattare con la pubblica amministrazione per un certo numero di mesi – si pensi, ad esempio, a una società operante nello sviluppo di sistemi di intelligenza artificiale destinati proprio a enti pubblici – ovvero il divieto di pubblicizzare beni o servizi per pari o superiore durata – particolarmente impattante per imprese che "vivono", fondandosi il proprio successo commerciale, di pubblicità e di *marketing* e/o che propongono ai propri clienti soluzioni innovative digitali basate anche su sistemi di intelligenza artificiale (come, ad esempio, nel settore assicurativo o della *cybersecurity*).

In tale contesto applicativo delle sanzioni interdittive, disposte in via cautelare interinale, nondimeno, trovano applicazione i limiti e le garanzie poste dal d.lgs. 231/2001, che sono in grado quanto meno di attenuarne portata ed effetti.

Tra gli altri, si menzionano i presupposti applicativi previsti dall'articolo 13, comma 1 d.lgs. 231/2001, in forza dei quali non è sufficiente la mera comminatoria legale delle sanzioni interdittive (come quella contenuta nell'articolo 25-*vicies* d.lgs. 231/2001), ma occorre la sussistenza di ulteriori condizioni, rappresentate, alternativamente, dalla sussistenza di «*un profitto di rilevante entità*» che l'ente ha tratto dal reato e dalla circostanza che «*il reato è stato commesso da soggetti in posizione apicale ovvero da soggetti sottoposti all'altrui direzione quando, in questo caso, la commissione del reato è stata determinata o agevolata da gravi carenze organizzative*» (lett. a) ovvero dalla «*reiterazione degli illeciti*» (lett. b), con rinvio implicito alla disposizione dell'articolo 20 d.lgs. 231/2001).

Lo stesso articolo 13 d.lgs. 231/2001, nel comma 2, limita l'estensione diacronica della sanzione interdittiva (anche in fase cautelare) nel *range* temporale compreso tra tre mesi e due anni, impedendo l'applicazione *ad libitum iudicis* della durata dell'interdizione.

Infine, occorre rammentare che, alla luce della disciplina dell'articolo 46 d.lgs. 231/2001, il giudice penale dovrà tarare la cautela interdittiva tenendo conto – analogamente a quanto previsto per le persone fisiche (articolo 275 c.p.p.) – di una serie di principi fondamentale in tema di «*Criteri di scelta delle misure*».

Il giudice competente, in sede di emissione dell'ordinanza cautelare, dovrà considerare, in primo luogo, il principio di adeguatezza della misura, inteso quale «*specificità idoneità di ciascuna in relazione alla natura e al grado delle esigenze cautelari da soddisfare nel caso concreto*», il quale gli impone, una volta sciolto l'*an* circa il ricorso allo strumento cautelare, di applicare la misura non "alla cieca", indiscriminatamente rispetto all'attività aziendale nel suo complesso, bensì in relazione a quella sola fase, processo o "porzione" di *business* (ad esempio, l'area di sviluppo di servizi digitali per i clienti) che ha

colposamente agevolato la commissione dell'illecito da parte dell'ente e rispetto alla quale si pone come *«concreto il pericolo che vengano commessi illeciti della stessa indole di quello per cui si procede»*, in relazione anche alle caratteristiche specifiche della violazione del caso di specie (riferibile, in senso lato, alla gestione endoaziendale dei sistemi di intelligenza artificiale).

Sotto questo profilo, inoltre, il giudice dovrà garantire il rispetto del principio di proporzionalità della misura, affinché essa sia, giustappunto, *«proporzionata all'entità del fatto e alla sanzione che si ritiene possa essere applicata all'ente»*, in modo da evitare l'applicazione di provvedimenti interdittivi oltremodo gravosi per l'organizzazione, anzitutto dal punto di vista della latitudine applicativa della misura in relazione all'attività aziendale. Esclusa la rilevanza del principio di sussidiarietà (valevole per l'interdizione dall'esercizio dell'attività, come visto non applicabile all'ente per espressa scelta legislativa), da ultimo, il giudice penale dovrà tenere a mente il principio di esclusività della cautela interdittiva, dal momento che le misure – tra quelle previste in guisa di sanzioni dall'articolo 25-*vicies*, comma 3 d.lgs. 231/2001 – *«non possono essere applicate congiuntamente»*.

Il perimetro di responsabilità dell'ente è stato ulteriormente esteso al delitto di *«Illecita diffusione di contenuti generati o alterati con sistemi di intelligenza artificiale»* punito dall'articolo 612-*quater* c.p. – disposizione già in precedenza introdotta dalla l. 132/2025 – in relazione al quale il trattamento sanzionatorio della *societas* risulta articolato su di una sanzione pecuniaria compresa tra duecento e settecento quote e sulla potenziale irrogazione delle sanzioni interdittive previste dall'articolo 9, comma 2, lett. b), c), d) ed e) d.lgs. 231/2001.

Si rammenta – rinviando per il dettaglio agli approfondimenti compiuti dai commentatori della l. 132/2025¹⁰ – che il delitto dell'articolo 612-*quater* c.p. costituisce reato comune (non richiedendo specifiche qualifiche giuridiche o di fatto dell'agente), inserito tra i delitti contro la libertà morale della persona, finalizzato a colpire con lo stigma della sanzione penale la realizzazione e la diffusione di c.d. *deepfakes*, ossia di contenuti multimediali che riproducono falsamente persone, oggetti e luoghi reali e li manipolano in modo ingannatorio, da cui derivi un danno ingiusto (sia esso patrimoniale che reputazionale).

5.2. Il “catalogo” dei reati presupposto: le modifiche indirette.

Occorre considerare, altresì, che la stessa l. 132/2025 aveva già apportato un'altra modifica di natura “indiretta” al d.lgs. 231/2001, introducendo una serie di circostanze aggravanti speciali – per il caso in cui *«il fatto è commesso mediante l'impiego di sistemi di intelligenza artificiale»* – nelle fattispecie di

¹⁰ Per un approfondimento sul delitto dell'articolo 612-*quater* c.p., si rinvia a A. MILONE, *Il diritto penale alla prova dell'intelligenza artificiale: alcune riflessioni intorno all'art. 612-*quater* c.p.*, in *Giur. pen. web*, 2026, 5.

aggiotaggio societario (articolo 2637, comma 2 c.c.), costituente reato presupposto della responsabilità dell'ente ai sensi dell'articolo 25-ter, comma 1, lett. r) d.lgs. 231/2001 («*Reati societari*») e di manipolazione del mercato (articolo 185, comma 1, secondo periodo d.lgs. 58/1998), costituente reato presupposto della responsabilità dell'ente ai sensi dell'articolo 25-sexies d.lgs. 231/2001 («*Abusi di mercato*»).

Se per tali ipotesi criminose l'estensione del perimetro di rilevanza delle condotte anche nella prospettiva del d.lgs. 231/2001 discende dal loro richiamo all'interno del "catalogo" dei reati presupposto, una seconda modifica apportata dalla L. 132/2025, di natura parimenti "indiretta", si caratterizza per la propria portata ubiquitaria e trasversale rispetto a tutte le fattispecie di reato presupposto contemplate dagli articoli 24 ss. d.lgs. 231/2001.

Infatti, l'introduzione nella "parte generale" del Codice penale di una circostanza aggravante, relativa all'aver «*commesso il fatto mediante l'impiego di sistemi di intelligenza artificiale, quando gli stessi, per la loro natura o per le modalità di utilizzo, abbiano costituito mezzo insidioso, ovvero quando il loro impiego abbia comunque ostacolato la pubblica o la privata difesa, ovvero aggravato le conseguenze del reato*» (articolo 61, comma 1, n. 11-decies c.p.) – costituente circostanza "comune", come tale applicabile a tutti i reati – esplica effetti, nella prospettiva del d.lgs. 231/2001, sulle modalità di commissione (e sulla relativa identificazione) di tutti i reati contemplati dal "catalogo".

In questo senso, pertanto, tutte le condotte di reato ricomprese nel novero dei reati presupposto – ad eccezione, ragionevolmente, di quelle ipotesi che, in concreto, difficilmente possono realizzarsi «*mediante l'impiego di sistemi di intelligenza artificiale*» (si pensi, ad esempio, alle pratiche di mutilazione degli organi genitali femminili dell'articolo 583-bis c.p., previste dall'articolo 25-quater.1 d.lgs. 231/2001, o ai delitti contro gli animali contemplati dall'articolo 25-undecies d.lgs. 231/2001) – risultano "sensibili" rispetto alle modalità di commissione descritte nella nuova aggravante e, dunque, alla realizzazione del fatto di reato in presenza vuoi di una situazione di "minorata difesa" (e, per converso, dalla prospettiva del soggetto agente, di "maggiorata insidiosità" della condotta) vuoi di oggetto aggravamento del reato derivanti dall'uso di sistemi di intelligenza artificiale.

Si noti, tuttavia, che l'aggravante in esame, dal punto di vista del richiamato rischio "da intelligenza artificiale" esplica una portata alquanto ampia, poiché – a differenza delle fattispecie criminose previste dall'articolo 437-bis c.p., contemplato dall'articolo 25-vicies d.lgs. 231/2001 – riguarda tutti i sistemi di intelligenza artificiale impiegati nell'organizzazione, indipendentemente dal loro livello di "rischio", e non soltanto i sistemi a rischio alto.

Inoltre, tale rischio "da intelligenza artificiale" non è sovrapponibile a quello prefigurato dall'articolo 25-vicies d.lgs. 231/2001, giacché quest'ultimo (in

relazione alle ipotesi omissive dell'articolo 437-*bis*, commi 1 e 3 c.p.) riguarda solo – secondo la tesi sostenuta – specifici operatori della filiera dell'intelligenza artificiale, mentre il primo, ove il termine «*impiego*» fosse inteso in senso restrittivo come sinonimo di "utilizzo", colpirebbe, a rigore, soltanto l'ente che agisce come *deployer* di un sistema di intelligenza artificiale utilizzato sotto la propria autorità e per finalità esclusivamente professionali.

Non si ignora, peraltro, che l'aggravante in esame pare pacificamente applicabile – proprio in ragione del ricorso a un termine neutro ed estensivo come "impiego" – anche rispetto a fattispecie criminose in cui il sistema di intelligenza artificiale «*sia utilizzato nel corso di un'attività personale non professionale*», le quali, pertanto, alla luce del comparto definitorio dell'AI Act, esulano sia dal concetto di *deployer* che, in ogni caso, dal perimetro degli operatori della filiera dell'intelligenza artificiale.

In queste ipotesi, in particolare, potrebbe porsi il quesito se l'ente – per il tramite di una sua persona fisica qualificata che "impiega" un sistema di intelligenza artificiale "privato" o, comunque, "aziendale" ma al di fuori del contesto professionale – possa comunque essere chiamato a rispondere ai sensi del d.lgs. 231/2001, per un reato presupposto aggravato poiché commesso «*mediante l'impiego di sistemi di intelligenza artificiale*» (con le specifiche modalità previste dall'articolo 61 c.p.), nella misura in cui quel reato, benché realizzato attraverso l'utilizzo di un sistema di intelligenza artificiale, ma al di fuori di un "ambiente" e/o di finalità professionali, abbia concretizzato un interesse dell'ente (*ex ante*) o, molto più probabilmente, abbia determinato in suo favore un vantaggio (da valutare *ex post* e in chiave oggettiva).

Si intersecano con la fattispecie applicativa in esame la problematica nota come c.d. *shadow AI*, espressione con cui si fa generalmente riferimento – seppur in via non esclusiva – all'utilizzo non autorizzato (e talora non tracciato o tracciabile dall'organizzazione), da parte del personale aziendale, di sistemi di intelligenza artificiale generativa per finalità inerenti l'esercizio della propria attività professionale, in chiave di miglioramento della produttività e di velocizzazione dei processi.

Si pensi, in tal senso, alle (invero piuttosto diffuse) pratiche di *upload*, su applicativi di intelligenza artificiale *open source* – esclusi dalla "*white list*" della società – di documentazione aziendale, talora anche "riservata" e "sensibile" (ad esempio, perché recante dati personali di clienti o fornitori, magari "particolari" o giudiziari; coperta dal segreto industriale; contenente informazioni privilegiate *insider-sensitive*), per finalità di analisi e di elaborazione di dati e/o di redazione e revisione di informative, contratti, *template* o pareri legali a uso interno.

Tanto premesso, deve essere brevemente analizzato il significato semantico della circostanza aggravante in esame, la quale – come visto – può essere idealmente scomposta in tre diverse fattispecie.

La prima fattispecie, che riguarda l'ipotesi in cui i sistemi di intelligenza artificiale impiegati per commettere il fatto *«per la loro natura o per le modalità di utilizzo, abbiano costituito mezzo insidioso»*, fa perno sull'effetto dell'utilizzo del sistema (ossia la sua caratteristica di insidiosità) sulla persona offesa o danneggiata e può scaturire, anzitutto, dalle sue concrete *«modalità di utilizzo»*, formula che, giustapposta alla nota della "insidiosità", sembra richiamare indirettamente la circostanza aggravante del "mezzo fraudolento" già prevista in materia di furto (articolo 625, comma 1, n. 2 c.p.), la quale, secondo la corrente giurisprudenza, ricorre appunto quando l'agente utilizza una particolare astuzia, stratagemma o insidia per superare la naturale vigilanza della vittima e rendere più facile l'impossessamento della cosa altrui.

L'altra ipotesi di "insidiosità" del sistema di intelligenza artificiale, correlata alla sua *«natura»*, appare invero al limite del pleonasma, dal momento che potrebbe agevolmente sostenersi che tutti i sistemi di intelligenza artificiale, in ragione delle loro caratteristiche progettuali e di funzionamento (si pensi, tra le altre, alla nota tematica della c.d. *black box*), sono per *«natura»* intrinsecamente insidiosi.

Se così interpretata, l'aggravante assumerebbe un'indubbia coloritura oggettivistica – che la renderebbe pressoché sempre contestabile – di dubbia compatibilità costituzionale, in relazione al principio di colpevolezza-personalità della responsabilità penale.

La seconda fattispecie di aggravante correlata all'utilizzo di sistemi di intelligenza artificiale ricorre *«quando il loro impiego abbia comunque ostacolato la pubblica o la privata difesa»*, formula che richiama in modo esplicito la circostanza aggravante della "minorata difesa" (articolo 61, comma 1, n. 5 c.p.), la quale – nel contesto in esame – può ritenersi applicabile al caso in cui, ad esempio, non è fornita adeguata trasparenza all'utente finale che egli sta interagendo con un operatore "virtuale" (quale un *voice bot*) e ciò consenta di perpetrare ai suoi danni una truffa o un reato informatico o in materia di *privacy*.

L'ultima fattispecie ricorre allorché l'impiego del sistema di intelligenza artificiale abbia *«aggravato le conseguenze del reato»*, circostanza che potrebbe rinvenirsi nelle ipotesi in cui l'utilizzo del sistema (si pensi a un caso di riciclaggio o di autoriciclaggio) abbia reso estremamente difficile, se non impossibile, ricostruire il *paper trail* o, comunque, la destinazione e i successivi utilizzi di eventuali proventi del reato.

L'inserimento della nuova aggravante codicistica comporta l'interpolazione nelle pieghe del "catalogo", in via surrettizia, del rischio reato "da intelligenza artificiale", con conseguente necessità di "diffusione" dei presidi di

mitigazione del rischio reato in relazione alle varie – in molti casi non mappate o, comunque, non immediatamente tracciate – attività aziendali in cui può verificarsi un utilizzo di sistemi di intelligenza artificiale.

Si allude, naturalmente – almeno per quanto riguarda i reati presupposto dolosi previsti dal “catalogo” del d.lgs. 231/2001 – a un impiego “strumentale” dei sistemi di intelligenza artificiale, ossia teleologicamente preordinato, in via diretta o indiretta, alla realizzazione di un reato presupposto o, comunque, di condotte preparatorie e funzionali alla successiva commissione dell’illecito (non necessariamente costituenti reato, ma magari semplicemente illecito sanzionabile in via amministrativa ai sensi dell’articolo 99 Al Act e del d.lgs. 179/2026).

Per quanto concerne i reati colposi previsti dal “catalogo”, l’analisi è senz’altro più complessa, dal momento che – almeno dal punto di vista letterale – l’aggravante sembrerebbe attribuire rilievo al dato oggettivo, in sé e per sé, della commissione del fatto mediante l’impiego di sistemi di intelligenza artificiale (secondo le modalità specificate nell’ articolo 61, comma 1, n. 11-*decies* c.p.), non soggettivamente connotato in termine di volontà di “abusare” dello strumento per conseguire i risultati divisati dall’aggravante. In effetti, la disposizione che reca l’aggravante – proprio perché qualificabile come forma di “minorata difesa” da uso di sistemi di intelligenza artificiale – appare “sbilanciata” sulla posizione della persona offesa o danneggiata dal reato e sugli effetti che il fatto ha prodotto sulla sua sfera, in termini di insidiosità-offensività o di ostacolo alla difesa, nonché in relazione all’obiettivo aggravamento delle conseguenze del reato.

In altri termini, l’aggravante parrebbe poter assumer una connotazione oggettiva, di sicura compatibilità con l’agire meramente colposo (negligente, imprudente, imperito o inosservante di regolare cautelari specifiche, quali quelle dettate dall’Al Act) della persona fisica che, nell’esercizio della propria attività, impiega il sistema di intelligenza artificiale e, senza rappresentarsi e volere l’evento offensivo, realizza un fatto di reato.

Un’interpretazione “spinta” – di dubbia tenuta concettuale – potrebbe forse mutuare i principi elaborati dalla giurisprudenza sul d.lgs. 231/2001 in materia di reati colposi di evento – che, com’è noto, afferma la compatibilità tra i requisiti dell’interesse e del vantaggio e la mancanza di volontà tipica del reato colposo, imponendo di correlare i due criteri di imputazione oggettiva alla condotta inosservante e non all’evento lesivo finale – estendendoli e adattandoli al diverso contesto di analisi in discorso, per ritenere comunque, anche volendo attribuire una sfumatura soggettiva all’aggravante in discorso, pienamente compatibile l’aggravante con la natura colposa del fatto, sul presupposto della riferibilità della consapevolezza circa l’impiego “distorto” del sistema di intelligenza artificiale (ad esempio, per insidiare o per superare la “difesa” del soggetto offeso) alla sola fase antecedente della condotta e non a quella dell’evento finale.

In questo senso, dunque, anche i reati colposi oggi già previsti dal “catalogo” – il riferimento è alle fattispecie richiamate dagli articoli 25-*septies* (omicidio colposo e lesioni personali colpose, gravi e gravissime) e 25-*undecies* d.lgs. 231/2001 (reati ambientali) – potrebbero astrattamente essere connotati dall’aggravante in questione.

Volendo ipotizzare un caso di scuola, potrebbe incolparsi l’ente, in qualità di *deployer*, per un infortunio sul lavoro determinato dall’impiego (imputabile al datore di lavoro e, a seconda dei casi, agli altri garanti in materia prevenzionistica) di una macchina o di un’attrezzatura a funzionamento autonomo o semiautonomo (quali un *robot* o un *cobot*), basato su sistemi di intelligenza artificiale, qualora la messa a disposizione dei lavoratori non sia stata consapevolmente preceduta – per evitarne i relativi costi e tempi – da un’adeguata formazione, da una dettagliata informazione e da un addestramento specifico, con la conseguenza che l’infortunio si è determinato (o si è aggravato) per l’incapacità del lavoratore, in termini di sua “minorata difesa”, di interagire correttamente con il macchinario o, comunque, per l’oggettiva “insidiosità” derivante dalla natura *AI-based* dello stesso.

5.3. Linee di intervento per la mitigazione del rischio reato “da intelligenza artificiale”.

L’ampliamento e la conseguente “apertura” del “catalogo” dei reati presupposto ai rischi – diretti (articolo 25-*vicies* d.lgs. 231/2001) e indiretti (aggravanti comuni e speciali) – “da intelligenza artificiale” chiamerà le imprese, attraverso le proprie funzioni interne di *compliance* (e, ove non presenti, con il supporto di consulenti esterni), a una significativa attività di aggiornamento del modello di organizzazione, gestione e controllo dell’ente. L’operazione logicamente preliminare consisterà nell’attività di analisi e di identificazione dei (nuovi) rischi reato nel contesto aziendale di riferimento, funzionale – al di fuori del caso in cui i rischi appaiano non pertinenti rispetto a essa – all’individuazione dei processi e delle attività “sensibili”, alla valutazione del rischio e alla successiva definizione, nell’ambito della strategia aziendale di trattamento del rischio, dei relativi protocolli di prevenzione (*rectius*, di mitigazione) dei suddetti rischi reato.

Al fine di svolgere la *risk analysis* e il successivo *risk assessment*, le organizzazioni potranno valorizzare, quale utile “sponda”, le attività di mappatura, di censimento e di classificazione dei sistemi di intelligenza artificiale eventualmente già intraprese – o che avrebbero dovuto intraprendere – in vista delle scadenze applicative definite dall’AI Act. Dette attività, infatti, costituiscono, in prima battuta, in una prospettiva di *compliance* regolatoria, il presupposto necessario per identificare i livelli di rischio dei sistemi o modelli di intelligenza artificiale a vario titolo sviluppati o utilizzati, nonché in seconda battuta, nella chiave preventiva del d.lgs.

231/2001, per acquisire una conoscenza strutturata della realtà aziendale in termini di “ecosistema” dell’intelligenza artificiale.

In questa duplice prospettiva, l’attività di mappatura consente di identificare i sistemi di intelligenza artificiale, distinguendoli in funzione della relativa qualificazione normativa (pratiche «vietate», ad alto rischio, a rischio limitato ovvero a rischio minimo), della tipologia tecnologica adottata (*machine learning*, *deep learning*, AI generativa, *natural language processing*, *computer vision*) e dell’eventuale utilizzo di modelli di intelligenza artificiale per finalità generali (GPAI), e, al contempo, di individuare i processi e sotto-processi (o attività) aziendali nei quali tali sistemi risultano integrati (in via meramente esemplificativa, HR, *procurement*, *Third Party Management*, *cybersecurity*).

La mappatura dei sistemi di intelligenza artificiale e dei processi aziendali a essi correlati consente di individuare, in una seconda fase, le funzioni e i soggetti coinvolti nelle diverse fasi del “ciclo di vita” del sistema – a seconda del ruolo che l’ente può ricoprire: sviluppo, acquisizione, personalizzazione (*fine tuning*, *retraining*) e utilizzo – e di comprendere come effettivamente si svolgono i percorsi decisionali e i flussi informativi nell’ambito del processo in cui è “innestato” il sistema di intelligenza artificiale.

Collaterale – e, per certi versi, logicamente prioritaria – rispetto a tale attività di mappatura dei sistemi di intelligenza artificiale è la definizione (o, ove già presente, la ricostruzione) della *governance* dell’intelligenza artificiale, volta a identificare ruoli e poteri all’interno dell’organizzazione per garantire la “conformità” all’AI Act e alle disposizioni attuative (normative e di carattere tecnico), ciò anche in una prospettiva di attribuzione delle responsabilità “all’interno” dell’ente e “verso l’esterno”.

Quest’ultimo profilo, nel dettaglio, chiama direttamente in causa l’allocazione del rischio sanzionatorio, amministrativo e penale, con riferimento agli illeciti correlati al rispetto della disciplina in materia di intelligenza artificiale, ambito nel quale, tuttavia – come ad esempio accade anche nel settore della *privacy* – non esiste una tipizzazione a livello normativo delle catena interna di soggetti “garanti”, a cui può essere imputata la relativa responsabilità nell’ambito dell’organizzazione, ma soltanto la definizione di macro-soggetti (*deployer*, *provider* e altri) titolari dei rispetti obblighi regolatori.

Premesso, dunque, che le responsabilità per le violazioni delle disposizioni dell’AI Act (come sanzionate in base alla disciplina nazionale di adeguamento) faranno capo all’organizzazione nel proprio complesso in qualità, ad esempio, di *provider* o di *deployer*, tuttavia, dal momento che gli illeciti amministrativi e penali si fondano sul principio di personalità della responsabilità, si pone il tema – noto ed esplorato da tempo in altri settori del diritto d’impresa (salute e sicurezza sui luoghi di lavoro, ambiente, tributario, *privacy*) – di valutare l’ammissibilità (anzitutto dal punto di vista giuridico) e la concreta fattibilità del conferimento di “deleghe” opponibili ai

terzi, aventi l'effetto di ripartire gli obblighi in materia di intelligenza artificiale e le connesse responsabilità.

Rifacendosi alla corrente elaborazione giurisprudenziale, può ipotizzarsi il conferimento, anzitutto, di una delega "di attribuzioni" o gestoria, ai sensi dell'articolo 2381 c.c., nell'ambito dell'organo amministrativo collegiale in favore di uno o più dei suoi membri, con l'effetto di "concentrazione" delle responsabilità in capo al/i delegato/i, pur sempre al cospetto di un persistente obbligo di vigilanza "alta" in capo agli altri membri dell'organo amministrativo sulla generale gestione del rischio da parte del/i delegato/i. Sia conferita o meno una delega gestoria, maggiormente controvertibile – in assenza di un'espressa previsione normativa in tal senso (sulla falsariga dell'istituto previsto dall'articolo 16 d.lgs. 81/2008 in materia di salute e sicurezza sul lavoro, poi esteso in via interpretativa ad altri settori come la tutela dell'ambiente, la sicurezza alimentare e la responsabilità previdenziale) – è il conferimento di una delega "di funzioni" all'esterno dell'organo amministrativo, con la finalità di "traslare" gli obblighi dal garante originario (membri dell'organo amministrativo o delegato/i ai sensi dell'articolo 2381 c.c.) verso un soggetto delegato, così chiamato a rispondere delle rispettive violazioni, dal momento che né l'AI Act, né la normativa nazionale di adeguamento prevedono tale istituto giuridico.

La prospettiva organizzativa, in altri termini, sarebbe quella di conferire, con apposita procura speciale, una delega "di funzioni" a un soggetto dotato di specifica esperienza e competenza professionale in materia di intelligenza artificiale, di poteri di organizzazione, gestione e controllo, nonché di correlati e adeguati poteri di spesa, il quale potrebbe essere identificato in un ipotetico "*AI manager*" che dovesse essere nominato da parte dell'ente, quale figura aziendale parimenti non contemplata dalla normativa regolatoria, da mantenere idealmente distinta dall'"*AI Officer*" (soggetto accostabile, per ruolo e prerogative, al *Data Protection Officer*, seppur, diversamente da quest'ultimo, non disciplinato dalla normativa europea e nazionale).

In assenza di specifiche indicazioni sul punto, peraltro, allo stato attuale il rischio è che l'eventuale conferimento di una simile delega a un soggetto interno all'organizzazione possa essere disconosciuta ai fini sanzionatori da parte delle autorità competenti (regolatorie e giudiziarie penali), ritenendo che tale delega abbia una portata esclusivamente interna (assimilabile piuttosto a un incarico di esecuzione, privo di effetto devolutivo) tale da lasciare intatta la posizione di garanzia originaria dei membri dell'organo amministrativo (o dell'eventuale soggetto destinatario della delega gestoria) e, in definitiva, il suo rilascio deve essere attentamente valutato e compreso nella sua effettiva portata.

Si osserva come, in ogni caso, dalla prospettiva della responsabilità "da reato" degli enti, l'eventuale incertezza circa l'identificazione, all'interno

dell'organizzazione, del/i soggetto/i autore/i del reato presupposto e, comunque, lungo la scala direzionale, tenuto/i all'adempimento "verso l'esterno" degli obblighi regolatori in materia di intelligenza artificiale – se potrebbe porre ostacoli all'accertamento di una responsabilità penale individuale «*al di là di ogni ragionevole dubbio*» – non osta al riconoscimento della responsabilità della *societas*, giacché – com'è noto – l'articolo 8 d.lgs. 231/2001 pone quale antidoto alla possibile "dispersione" delle responsabilità nelle organizzazioni complesse il principio di autonomia, in forza del quale l'ente risponde del reato anche quando «*l'autore del reato non è stato identificato*» (comma 1, lett. a)).

Le sopra descritte attività di "mappatura" (dei sistemi di intelligenza artificiale e della *governance*) rappresentano un passaggio obbligato in fase di aggiornamento del modello organizzativo previsto dal d.lgs. 231/2001, nella prospettiva di analisi, di valutazione e di mitigazione tanto dei rischi "diretti" – correlati all'omessa adozione di misure tecniche di sicurezza o di sorveglianza umana (nonché di alterazione) dei sistemi di intelligenza artificiale ad alto rischio (articolo 437-*bis* c.p.) – quanto dei rischi "indiretti" – di carattere ubiquitario, legati al mero "impiego" di un qualunque sistema di intelligenza artificiale nella realizzazione di una condotta illecita prevista dal "catalogo" dei reati presupposto.

In tale prospettiva "strumentale", la ricognizione dei sistemi di intelligenza artificiale, a seconda dei casi, sviluppati e/o utilizzati nel contesto aziendale e la "fotografia" della *governance* aziendale in materia dovrebbe affiancare e integrare l'analisi specificamente orientata all'individuazione delle aree di esposizione ai nuovi rischi reato, fungendo da "termometro" per la definizione del livello di rischio reato (inerente, attuale e residuo) e da filtro selettivo per la corretta identificazione delle misure di prevenzione, organizzative e operative, che l'ente è chiamato ad adottare e a implementare.

La fase successiva, relativa alla definizione delle misure di prevenzione, implica l'adozione (nel caso l'ente ne sia sfornito del tutto) ovvero l'aggiornamento dei protocolli organizzativi e dei controlli interni, secondo un approccio "integrato", atto a valorizzare le sinergie che – inevitabilmente, data la portata giuridica trasversale delle tematiche dell'intelligenza artificiale – si innestano tra quest'ultima e altre normative regolatorie già presidiate in ambito aziendale e altrettanto rilevanti nella prospettiva di rischio del d.lgs. 231/2001, quali, tra le altre, la protezione dei dati personali, la *cybersecurity*, la tutela della proprietà industriale, le normative *antitrust* e sul *market abuse*. Tra i principali protocolli organizzativi che richiedono – data la loro rilevanza quale presidio del rischio reato – implementazione immediata (o, quanto meno, non eccessivamente procrastinata nel tempo) nell'organizzazione aziendale sono destinati ad assumere portata prioritaria:

- l'adozione di una *governance* aziendale dell'intelligenza artificiale che disciplini, nell'ambito di un organigramma strutturato e mediante il conferimento di deleghe e incarichi, ruoli, responsabilità e processi decisionali connessi allo sviluppo e/o all'utilizzo di sistemi di intelligenza artificiale;
- l'aggiornamento del codice etico attraverso l'inserimento dei principi – come risultanti dall'AI Act e dalla normativa di adeguamento – che presiedono alle varie fasi del "ciclo di vita" dei sistemi di intelligenza artificiale;
- l'adozione – a livello locale e di gruppo, a seconda delle dimensioni e delle articolazioni territoriali dell'ente – di *policy*, procedure e istruzioni operative preordinate a delineare la *governance* aziendale dell'intelligenza artificiale e a definire il riparto di attribuzioni nell'ambito della c.d. *process value chain*, nonché a disciplinare, nelle varie fasi di cui si compongono, lo sviluppo, l'addestramento, l'immissione sul mercato, la messa in servizio, la messa a disposizione sul mercato e l'utilizzo dei sistemi di intelligenza artificiale, nonché le varie "vicende" cui essi possono andare incontro nel proprio "ciclo di vita" (in termini di monitoraggio, gestione dei rischi, riesame, gestione delle "emergenze");
- l'introduzione di processi di approvazione e autorizzazione preventive per l'utilizzo di modelli generativi «a rischio sistemico» e di sistemi di intelligenza artificiale ad alto rischio;
- la regolamentazione dell'impiego di sistemi di intelligenza artificiale da parte del personale, privilegiando, ove possibile, applicativi governati nell'ambito dell'infrastruttura informatica aziendale e limitando il ricorso a sistemi *open source*, rispetto ai quali l'ente non è in grado di esercitare un controllo effettivo in termini di dati di *input* e di *output*; nonché il contestuale divieto di utilizzo di sistemi di intelligenza artificiale "personali" per finalità professionali, non mappati, in chiave di prevenzione di fenomeni di *shadow AI*;
- l'introduzione di meccanismi di tracciabilità delle operazioni compiute attraverso l'utilizzo di sistemi di intelligenza artificiale;
- l'introduzione di misure di supervisione umana effettiva nei processi "sensibili", accompagnate da un'adeguata documentazione delle attività di controllo svolte;
- l'adeguamento della documentazione contrattuale, mediante l'inserimento di clausole – da modulare, caso per caso, a seconda del ruolo assunto dall'ente lungo la catena dell'intelligenza artificiale, del livello di rischio del sistema e della tipologia di sistema (ad esempio, un GPAI) – che, tra le altre, vincolino il *deployer* all'uso di un sistema ad alto rischio nel rispetto delle istruzioni (ivi inclusa la finalità del sistema) e della documentazione tecnica predisposta dal *provider*

ovvero che garantiscano il rispetto da parte del *provider* dei requisiti regolatori previsti per i medesimi sistemi;

- l'attivazione di programmi di formazione (*rectius*, alfabetizzazione) specialistica e di sensibilizzazione in favore dei soggetti maggiormente esposti ai rischi di commissione di reati connessi all'utilizzo di sistemi di intelligenza artificiale.

La costruzione di un'adeguata "rete" di controlli preventivi – che non può prescindere da una altrettanto adeguata e capillare attività di identificazione e di valutazione dei rischi, sui cui in definitiva si gioca la vera "personalizzazione" del modello organizzativo – costituisce (e costituirà sempre più nei prossimi mesi) un adempimento essenziale per gli enti, dalle multinazionali dall'articolata e sofisticata organizzazione alle imprese di matrice "familiare", nella prospettiva di ottenere – in caso di commissione di uno dei *predicate crime* connessi ai sistemi di intelligenza artificiale – l'esenzione da responsabilità ai sensi del d.lgs. 231/2001, attraverso la prova, per il tramite del modello organizzativo, dell'assenza di "colpa di organizzazione", intesa – com'è noto – quale *deficit* organizzativo che ha consentito o, comunque, reso più agevole la commissione del reato.

A tal fine, un importante supporto per le imprese nella definizione dei propri presidi preventivi – in linea con le più aggiornate tendenze giurisprudenziali che riconoscono un'efficacia rafforzata ai modelli organizzativi ispirati alle norme tecniche (oltre che a codici di comportamento e *best practice* di settore) – potrà derivare:

- dall'implementazione di un sistema di gestione, ancorché non necessariamente certificato, conforme alle specifiche della norma EN ISO/IEC 42001:2026 (*Information technology - Artificial intelligence - Management system*, AIMS, Già ISO/IEC 42001:2023) e della corrispondente versione italiana di recente aggiornata UNI CEI EN ISO/IEC 42001:2026 (Tecnologie informatiche - Intelligenza artificiale - Sistema di gestione);
- eventualmente integrato – per i soli *provider* di sistemi di intelligenza artificiale ad alto rischio, tenuti al rispetto dell'obbligo di legge relativo al «*sistema di gestione della qualità*» (articolo 17 AI Act) – con il recentissimo *standard* EN 18286:2026 (*Artificial intelligence - Quality management system for EU AI Act regulatory purposes*), sulla base dei quali potrà essere strutturato il più ampio sistema dei controlli interni con funzione preventiva degli inediti rischi reato "da intelligenza artificiale".

Conclusivamente, può affermarsi che l'ingresso dell'intelligenza artificiale nelle pieghe organizzative, gestionali e operative delle imprese ha determinato – a dispetto di chi affronta semplicisticamente il tema alla stregua di un mero *hype* transeunte, destinato a "sgonfiarsi" nel medio

termine – enormi opportunità in termini di sviluppo, ma al contempo pone alle imprese impegnative sfide in punto di obblighi di “conformità”, regolamentare e tecnica, che richiederanno alle funzioni competenti un approccio serio, effettivo e sostanziale ai relativi rischi, con ripudio di pratiche disinvolve di *cosmetic compliance*, espressione di poco lungimiranti politiche di *cost efficiency*, le quali, alla resa dei conti, sono suscettibili di generare, all’insegna di un “effetto *boomerang*”, pregiudizi (patrimoniali, reputazionali e sanzionatori) superiori rispetto agli sforzi ragionevolmente esigibili dagli operatori economici.