GIUSTIZIA INSIEME

Diritto e Innovazione

Verso un cambiamento della responsabilità algoritmica dell'ente

di Chiara Giovannini

26 maggio 2026

Sommario: 1. Scrutinare l'IA: standard probatori e bilanciamento con la proprietà intellettuale – 2. Lo scrutinio della decisione algoritmica – 3. Colpa di organizzazione tecnologica e delega algoritmica

1. Scrutinare l'IA: standard probatori e bilanciamento con la proprietà intellettuale

L'emersione dei sistemi di intelligenza artificiale nei processi decisionali dell'impresa impone una profonda revisione delle categorie tradizionali della responsabilità degli enti ai sensi del d.lgs. 8 giugno 2001, n. 231. La crescente delega di funzioni valutative, selettive e talora decisionali ai modelli algoritmici sollecita un interrogativo centrale in merito all'incidenza dell'affidamento di scelte rilevanti a sistemi automatizzati sulla configurazione della colpa di organizzazione e, correlativamente, sull'imputazione della responsabilità all'ente. Tale interrogativo non può essere affrontato attraverso la categoria, concettualmente impropria, della

responsabilità "dell'algoritmo", poiché il principio di colpevolezza impone di individuare nell'ente – e non nello strumento tecnologico – il centro di imputazione del deficit organizzativo. Sebbene l'algoritmo non sia soggetto dotato di doveri giuridici né centro autonomo di imputazione, la sua integrazione nei flussi decisionali aziendali lo rende componente strutturale dell'architettura organizzativa.

L'adozione di sistemi di intelligenza artificiale determina una trasformazione strutturale del giudizio di responsabilità da reato, ampliando l'orizzonte dell'accertamento verso una dimensione tecnico-scientifica. Il sindacato sulla colpa di organizzazione, tradizionalmente circoscritto alla verifica dell'adeguatezza dei modelli di gestione e controllo, deve oggi estendersi all'analisi del funzionamento interno degli apparati algoritmici adottati dall'ente. Ciò comporta l'esame dei dataset di addestramento, delle scelte di configurazione dei modelli, delle logiche inferenziali alla base degli output e delle modalità di supervisione umana previste. L'accertamento giudiziale richiede, pertanto, l'impiego di strumenti cognitivi propri delle scienze informatiche, quali perizie ad elevato contenuto tecnico, test retrospettivi di robustezza, simulazioni controfattuali e ricostruzioni del processo decisionale automatizzato, spesso fondato su modelli non lineari non riconducibili a strutture deterministiche tradizionali. In questa prospettiva, il processo 231 diviene luogo di intersezione tra sapere giuridico e competenze tecnologiche: pur non assumendo la veste di soggetto dell'illecito, l'algoritmo diventa oggetto necessario di scrutinio nella valutazione del deficit organizzativo. L'opacità tecnica non può, tuttavia, tradursi in una zona franca sottratta al controllo giudiziario, pena la compromissione del principio di effettività dell'accertamento e l'impossibilità di ricostruire il nesso eziologico tra assetto organizzativo e reato-presupposto.

Al contempo, l'impiego di sistemi proprietari o forniti da terzi introduce significative asimmetrie informative, idonee a incidere sull'effettività del contraddittorio tecnico. La limitata disponibilità del codice sorgente, dei parametri del modello o dei dataset di addestramento può impedire alle parti di verificare la correttezza del funzionamento del sistema, soprattutto quando il titolare del software invochi la tutela del segreto industriale. Ne deriva l'esigenza di un bilanciamento tra il diritto alla prova – che richiede la scrutinabilità del sistema rilevante ai fini dell'imputazione – e la protezione della proprietà intellettuale, che tutela interessi di rango non recessivo.

Tale bilanciamento trova oggi un primo ancoraggio sistematico nel Regolamento (UE) 2024/1689 (A.I. Act), il quale configura la riservatezza non come barriera assoluta, ma come criterio di modulazione dell'accesso alle informazioni. In particolare, l'art. 78 impone alle autorità coinvolte di garantire la tutela dei diritti di proprietà intellettuale e dei segreti commerciali, incluso il

codice sorgente, senza pregiudicare l'effettiva attuazione del Regolamento attraverso ispezioni, indagini e audit. Ne emerge un modello di "trasparenza funzionale", fondato su un accesso proporzionato e protetto alle sole informazioni strettamente necessarie alla valutazione del rischio, mediante soluzioni procedimentali quali l'accesso mediato tramite ausiliari tecnici, l'utilizzo di data-room protette o la limitazione dell'ostensione alle componenti rilevanti per l'accertamento.

Trasposta nel giudizio 231, tale logica impedisce che l'opacità algoritmica venga assolutizzata in chiave difensiva. La tutela del know-how, pur costituzionalmente e sovranazionalmente rilevante, non può tradursi in uno schermo idoneo a neutralizzare il diritto alla prova e l'effettività del controllo giudiziario. Il bilanciamento non può, pertanto, essere risolto in termini di alternativa secca tra *disclosure* e segretezza, ma richiede soluzioni intermedie capaci di garantire l'accesso funzionale alle informazioni necessarie all'accertamento senza compromettere in modo sproporzionato gli interessi proprietari dell'ente o del fornitore. In questa prospettiva, la tutela della proprietà intellettuale non si pone in antagonismo, ma si integra strutturalmente nel paradigma della responsabilità organizzativa, delimitando le modalità dell'ostensione senza escludere la verificabilità giuridica della scelta tecnologica e della sua governance.

2. Lo scrutinio della decisione algoritmica

L'evoluzione delle architetture algoritmiche, per quanto affascinante e in parte ispirata ai meccanismi del ragionamento umano, solleva rilevanti criticità sul piano giuridico, in particolare con riguardo alla possibilità di ricostruire il processo decisionale. I sistemi di intelligenza artificiale più avanzati, specie quelli fondati su tecniche non lineari di apprendimento automatico, operano infatti attraverso modelli che apprendono da grandi quantità di dati e producono output spesso refrattari a una piena spiegazione secondo le categorie deterministiche tradizionali. Tale opacità tecnica incide direttamente sulla responsabilità dell'ente, poiché nel giudizio ex d.lgs. 231/2001 il giudice è chiamato a ricostruire il percorso che ha condotto alla decisione rilevante ai fini del reato-presupposto.

In questo contesto, l'assenza di un'azione umana diretta e immediatamente verificabile impone all'ente un onere rafforzato di documentazione e tracciabilità del processo decisionale algoritmico. Occorre rendere ricostruibili i dati utilizzati, i criteri selezionati e ponderati, il margine di intervento umano e i controlli preventivi e successivi implementati, mediante la

conservazione di log delle decisioni automatizzate, protocolli di validazione, audit trail completi, report di testing e aggiornamento. In mancanza di tali presidi, la ricostruzione ex post del processo algoritmico rischia di essere impossibile, con conseguente qualificazione dell'opacità come indice sintomatico di una carenza organizzativa.

In tale prospettiva, la c.d. *explainability* assume un rilievo eminentemente probatorio: essa non si traduce nell'obbligo di una piena trasparenza matematica dei modelli complessi, ma richiede un livello minimo di intelligibilità funzionale che consenta il controllo giuridico della prevedibilità dell'evento, dell'evitabilità del reato e, in ultima analisi, della diligenza dell'assetto organizzativo. Non può essere tollerato che l'algoritmo si trasformi in una "scatola nera" rispetto alla quale l'ente non sia in grado di fornire spiegazioni plausibili e verificabili, pena l'indebolimento strutturale della difesa esimente.

L'accertamento dell'idoneità del modello organizzativo implica, inoltre, la verifica del nesso causale tra l'adozione o la gestione del sistema di IA e la commissione del reato-presupposto. Nel giudizio 231 è necessario distinguere tra causalità materiale, riferita alla condotta dell'autore umano del reato, e causalità organizzativa, riferita all'assetto dell'ente: l'algoritmo non è soggetto imputabile, ma strumento organizzativo che può costituire una condizione agevolatrice dell'illecito; l'indagine deve, quindi, accertare se il sistema sia stato adeguatamente validato, se eventuali bias o malfunzionamenti fossero conosciuti o conoscibili, se fossero previsti controlli umani effettivi e se le anomalie segnalate siano state ignorate o sottovalutate.

Il criterio di valutazione non è l'infallibilità del sistema, bensì l'adeguatezza dell'organizzazione rispetto al rischio conoscibile. Prevedibilità ed evitabilità assumono, in questo quadro, un ruolo centrale: qualora il reato derivi da errori tecnici segnalati dal fornitore, da dataset manifestamente distorti, dall'assenza di aggiornamento del modello o dalla mancata formazione del personale addetto alla supervisione, il deficit organizzativo assume un chiaro profilo colposo. Al contrario, ove l'anomalia sia tecnicamente imprevedibile secondo lo stato dell'arte, la rimproverabilità potrà risultare attenuata, a condizione che l'ente dimostri di aver adottato tutte le misure ragionevolmente esigibili per governare il rischio tecnologico. In questo scenario, l'accountability algoritmica – intesa come tracciabilità delle decisioni e documentabilità delle procedure – assume una funzione strutturale, rappresentando la condizione imprescindibile per la governabilità del rischio algoritmico e, in via derivata, per la stessa possibilità di esonero da responsabilità.

Essa svolge una triplice funzione: è criterio sostanziale di adeguatezza del modello organizzativo, parametro probatorio dell'idoneità e dell'effettiva attuazione del sistema di prevenzione, nonché

presidio preventivo capace di intercettare tempestivamente anomalie, bias o deviazioni rispetto agli scopi leciti. La sua rilevanza è particolarmente accentuata quando l'IA incide su aree sensibili ai reati-presupposto, rispetto alle quali l'ente deve dimostrare l'integrazione del modello in un circuito di controllo coerente con la mappatura dei rischi, attraverso la conservazione dei log di sistema, la documentazione dei criteri di addestramento, le verifiche periodiche delle performance e la possibilità di ricostruire la logica decisionale in termini funzionali.

Sotto il profilo sistematico, l'accountability tecnologica opera come criterio di delimitazione della rimproverabilità, evitando tanto che l'opacità algoritmica divenga un alibi organizzativo, quanto che l'errore tecnico si traduca automaticamente in responsabilità. Essa si inserisce nel più ampio principio di governabilità del rischio, secondo una logica di proporzionalità che impone presidi tanto più stringenti quanto maggiore è la complessità del sistema e la sua incidenza sui processi decisionali sensibili. In tale quadro, anche il ruolo dell'Organismo di Vigilanza risulta ridefinito: la sua funzione non può limitarsi a un controllo formale, ma deve estendersi alla verifica sostanziale del funzionamento dei sistemi algoritmici, mediante adeguate competenze tecniche interne o il ricorso a consulenze specialistiche. L'OdV deve monitorare report periodici, segnalazioni interne e gestione delle anomalie, poiché l'assenza di un controllo effettivo sull'uso dell'IA può integrare una carenza strutturale del sistema di vigilanza, idonea a incidere negativamente sul giudizio di idoneità complessiva del modello.

L'accountability tecnologica, infine, non costituisce un vincolo all'innovazione, ma il ponte che consente di integrare l'intelligenza artificiale in un sistema di prevenzione coerente con i principi di colpevolezza, prevedibilità ed esigibilità. L'ente non risponde per l'uso della tecnologia in quanto tale, bensì per l'incapacità di renderla trasparente, verificabile e governabile: è in questa capacità che oggi si misura la responsabilità organizzativa nell'era algoritmica.

3. Colpa di organizzazione tecnologica e delega algoritmica

L'introduzione di sistemi di intelligenza artificiale nei processi decisionali dell'impresa impone una riqualificazione della colpa di organizzazione alla luce della trasformazione qualitativa del rischio governabile. Nel paradigma del d.lgs. 8 giugno 2001, n. 231, la responsabilità dell'ente non discende dal solo verificarsi del reato-presupposto, ma dall'inadeguatezza strutturale dell'assetto organizzativo rispetto al rischio concretizzatosi: la colpa di organizzazione costituisce il fulcro dell'imputazione, quale giudizio di rimproverabilità fondato sulla prevedibilità ed evitabilità

dell'evento mediante un sistema diligente. L'adozione di architetture algoritmiche idonee a incidere in modo significativo sulle scelte dell'ente innalza lo standard di diligenza richiesto e ridefinisce i parametri di governo del rischio tecnologico.

In tale prospettiva si colloca la categoria della c.d. "colpa di organizzazione tecnologica", che non si traduce in un mero incremento quantitativo dei controlli, ma in un mutamento qualitativo del paradigma organizzativo necessario per fronteggiare rischi connotati da complessità, dinamicità e opacità. L'introduzione di sistemi predittivi non amplia automaticamente l'area della responsabilità né genera presunzioni di colpa, ma impone all'ente un onere di comprensione, presidio e supervisione delle logiche operative del modello, delle sue distorsioni potenziali e delle interazioni con i processi decisionali. L'innovazione tecnologica assume così la natura di atto organizzativo qualificato, non neutro, in quanto redistribuisce il rischio all'interno dell'apparato aziendale.

Ne deriva la necessità di un adattamento del modello 231, che da strumento di prevenzione della devianza individuale deve evolvere in dispositivo capace di governare processi decisionali ibridi, nei quali l'output è il risultato dell'interazione tra dati, algoritmi e supervisione umana. La diligenza organizzativa non può più misurarsi esclusivamente sulla linearità procedurale, ma sulla capacità effettiva di presidiare la complessità tecnologica introdotta. Tale ampliamento non implica responsabilità per mera complessità, poiché la colpa di organizzazione tecnologica resta ancorata ai criteri di prevedibilità ed evitabilità dell'evento secondo lo stato dell'arte tecnico e organizzativo: l'anomalia imprevedibile, la vulnerabilità non rilevabile o l'evento eccentrico rispetto alla mappatura dei rischi rimangono estranei all'area del rimprovero.

Sul piano operativo, la staticità dei presidi risulta incompatibile con la variabilità tecnica degli algoritmi. Il nuovo standard di diligenza richiede, pertanto, l'integrazione dei rischi algoritmici nell'analisi delle aree sensibili, nonché il rafforzamento dei presidi di controllo attraverso audit periodici, tracciabilità delle decisioni, formazione specialistica del personale e aggiornamento costante dei modelli, in ragione della loro naturale evolutività.

Sotto il profilo sistematico, la colpa di organizzazione tecnologica non altera la natura della responsabilità dell'ente, ma ne ridefinisce il contenuto sostanziale in relazione alla complessità tecnologica assunta dall'organizzazione. Il principio di colpevolezza conserva, in ogni caso, un ruolo centrale: l'ente risponde esclusivamente per deficit organizzativi causalmente rilevanti ed evitabili secondo la diligenza esigibile. Tuttavia, la complessità tecnologica introdotta dall'ente diviene parametro di valutazione dell'adeguatezza del governo del rischio: quanto maggiore è l'incidenza dell'algoritmo su aree sensibili ai reati-presupposto, tanto più elevato è lo standard di

governo richiesto. Tale ricostruzione impone, al contempo, un'applicazione proporzionata e modulata, idonea a evitare derive di vigilanza tecnologica permanente, eccessiva tecnicizzazione del giudizio penale o sproporzioni rispetto alle capacitàorganizzative delle imprese di minori dimensioni.

In questo quadro si inserisce la responsabilità da delega algoritmica, quale specifica declinazione della colpa di organizzazione nell'impresa tecnologicamente mediata. L'integrazione di sistemi di intelligenza artificiale in attività giuridicamente ed economicamente rilevanti non determina l'emersione di una responsabilità per il fatto della macchina né postula una soggettività giuridica dell'algoritmo, ma incide sulla struttura dell'assetto organizzativo cui resta imputabile il governo del rischio-reato. L'oggetto dell'imputazione non è l'autonomia tecnica del sistema, bensì la scelta dell'ente di affidare a strumenti algoritmici funzioni idonee a orientare la produzione, la distribuzione e il controllo del rischio, nonché l'adeguatezza dei presidi predisposti per disciplinarne selezione, impiego e supervisione.

Nel diritto penale dell'impresa, la delega costituisce un istituto strutturale della colpa di organizzazione: la ripartizione interna delle funzioni incide sui doveri di vigilanza e controllo senza determinare automatica esenzione di responsabilità. La delega algoritmica rappresenta una proiezione funzionale di tale schema in un contesto tecnologicamente mediato: il trasferimento non concerne poteri decisionali a un soggetto, ma l'incorporazione, nell'architettura organizzativa, di una discrezionalità tecnica cristallizzata in modelli predittivi e sistemi di decision-making automatizzato. È questa discrezionalità funzionale – e non l'autonomia ontologica dell'algoritmo, giuridicamente irrilevante – a ridefinire l'assetto decisionale dell'ente, il perimetro della supervisione umana e le modalità di presidio del rischio penalmente rilevante.

L'adozione di sistemi algoritmici in ambiti sensibili, quali procedure di gara, selezione dei partner commerciali, processi finanziari o controlli sulle transazioni, integra un atto organizzativo qualificato, idoneo a incidere in modo significativo sulla mappa dei rischi rilevanti ai fini 231. La responsabilità dell'ente discende dalla decisione di integrare tali strumenti nei processi aziendali e, soprattutto, dall'insufficienza delle regole che ne governano funzionamento, limiti operativi e controlli. In tal modo si evitano due esiti dogmaticamente distorti: da un lato, una responsabilità oggettiva tecnologica fondata sul mero errore imprevedibile dell'algoritmo; dall'altro, derive deresponsabilizzanti fondate sull'opacità o complessità del modello.

La responsabilità da delega algoritmica si articola attorno a tre nuclei essenziali di doveri organizzativi: un dovere di selezione consapevole, che impone una valutazione ex ante dei rischi

connessi all’impiego del sistema, dell’affidabilità del fornitore, delle logiche decisionali incorporate nel modello e dell’impatto del suo utilizzo sulle aree esposte ai reati-presupposto; un dovere di configurazione e integrazione, volto a inserire coerentemente l’algoritmo nei modelli di organizzazione, gestione e controllo ex d.lgs. 231/2001, mediante la definizione delle condizioni d’uso, delle soglie di intervento umano, delle responsabilità interne e dei flussi informativi verso l’Organismo di Vigilanza; un dovere di supervisione dinamica, che riconfigura il controllo umano attraverso monitoraggio continuativo delle performance, audit periodici, verifica di bias o anomalie, aggiornamento del modello e garanzia di tracciabilità delle decisioni automatizzate.

Nel giudizio di responsabilità ex d.lgs. 231/2001, la delega algoritmica rileva nella misura in cui il deficit di governance si ponga quale condizione agevolatrice del reato-presupposto. Il nesso causale non va ricercato in una relazione immediata tra output algoritmico ed evento illecito, bensì nel collegamento tra l’inadeguatezza dell’assetto decisionale e la concretizzazione del rischio. La responsabilità si configura, pertanto, come responsabilità per difetto di governo della delega tecnologica, entro i limiti della prevedibilità ed evitabilità dell’evento secondo lo standard di conoscenze tecniche e organizzative esigibili nel caso concreto.