5

AMBITO DI APPLICAZIONE E PRINCIPI GENERALI DELLA LEGGE 23 SETTEMBRE 2025, N. 132, RECANTE "DISPOSIZIONI IN MATERIA DI INTELLIGENZA ARTIFICIALE"[1]

Fortunato Costantino[2]

[1] **Como citar este artigo científico.** COSTANTINO, Fortunato. Ambito di applicazione e principi generali della legge 23 settembre 2025, n. 132, recante "disposizioni in materia di intelligenza artificiale. In: **Revista Amagis Jurídica**, Ed. Associação dos Magistrados Mineiros, Belo Horizonte, v. 18, n. 1, p. 143-191, jan.-abr. 2026.

[2] Professore Dottore in Giurisprudenza. Professore della European School of Economics. *E-mail*: costantinofortunato@libero.it

1 PREMESSA

La Legge 23 settembre 2025, n. 132, recante "*Disposizioni in materia di intelligenza artificiale*", rappresenta il primo intervento sistematico del legislatore italiano volto a disciplinare in modo trasversale l'impatto dell'intelligenza artificiale sulla società, sull'economia e sull'azione delle pubbliche istituzioni. Il provvedimento si colloca in una fase storica contrassegnata da una profonda accelerazione tecnologica, in cui i sistemi di IA, e in particolare quelli a carattere generativo o agentico, iniziano a incidere in modo strutturale su ambiti ad alta densità normativa e valoriale, quali la sanità, la giustizia, il lavoro, l'amministrazione pubblica, la formazione e la sicurezza.

La Legge n. 132/2025 si inserisce nel solco tracciato dal Regolamento (UE) 2024/1689 (c.d. AI Act), approvato dal Parlamento europeo nel marzo 2024 e destinato a entrare in vigore progressivamente entro il biennio successivo. In tale contesto, la scelta del legislatore nazionale non è stata quella di elaborare una disciplina autonoma, bensì di affiancare e integrare il quadro normativo europeo, in coerenza con l'articolo 117, primo comma, della Costituzione italiana e con il principio del primato del diritto dell'Unione. La legge italiana si propone così di fornire un inquadramento di principi e strumenti normativi utili a orientare l'interpretazione e l'applicazione dell'IA nel contesto ordinamentale interno, nel rispetto delle competenze nazionali e delle specificità del sistema giuridico italiano.

In questa prospettiva, i primi tre articoli della legge n. 132/2025 assumono una funzione fondativa e assiologica di sistema e delineano il perimetro di applicazione della disciplina, enunciano i valori ispiratori dell'intervento regolatorio e pongono le basi concettuali per la costruzione di un diritto dell'intelligenza artificiale coerente con i principi dello Stato costituzionale di diritto.

L'articolo 1 individua le finalità e l'ambito di applicazione della legge, collocandosi all'interno di un'impostazione esplicitamente antropocentrica e precauzionale dell'uso dell'IA, coerente con le raccomandazioni del Consiglio d'Europa e con i valori sanciti dalla Carta dei diritti fondamentali dell'Unione europea (Carta di Nizza).

L'articolo 2 fornisce la definizione tecnica e giuridica di "sistema" e "modello" di intelligenza artificiale", mutuando in larga parte la terminologia del Regolamento UE, e contribuendo a costruire la grammatica normativa della materia.

L'articolo 3, infine, enuncia una serie di principi generali – tra cui trasparenza, accuratezza, non discriminazione e sostenibilità – destinati a costituire criteri interpretativi e direttive per l'attuazione della legge da parte di tutti i soggetti coinvolti, pubblici e privati.

Letti nel loro insieme, tali articoli non si limitano semplicemente a introdurre una disciplina legale in ambito tecnologico, ma esprimono una dichiarazione di indirizzo politico, finalizzata a promuovere uno sviluppo tecnologico responsabile e sostenibile, che valorizzi l'innovazione nel rispetto dei diritti fondamentali, della legalità e della coesione sociale. In questo contesto, l'approccio normativo adottato si fonda su un equilibrio tra la stabilità giuridica dei principi generali e la flessibilità regolatoria necessaria per affrontare le mutazioni tecnologiche del futuro prossimo attraverso decreti attuativi, regolamenti e altre fonti secondarie. Queste fonti, unitamente alle future interpretazioni giurisprudenziali, rappresentano il meccanismo dinamico con cui la norma può rispondere alle esigenze di flessibilità e aggiornamento, senza compromettere i principi di base sanciti dalla legge.

L'intero impianto normativo prefigura insomma un modello di governance delle tecnologie emergenti che integra in modo armonico l'evoluzione della scienza e la protezione dei diritti fondamentali,

con una forte enfasi sull'accountability e sull'adattabilità delle norme. La legge, pur nella sua oggettiva necessità di ulteriori sviluppi normativi e interpretativi, è integralmente costruita su un principio di precauzione giuridica[3], volto a evitare che l'innovazione possa travolgere i diritti e i valori costituzionali su cui si fonda la democrazia italiana.

Nei paragrafi che seguono, si propone un'analisi sistematica dei primi tre articoli della legge, mediante il confronto tra il dato normativo e il quadro giuridico, nazionale ed europeo, vigente, con l'obiettivo di mettere in evidenza le principali implicazioni applicative e le criticità potenziali di una normativa destinata, per sua natura, ad accompagnare un'evoluzione tecnologica continua e dinamica. Particolare attenzione sarà dedicata al ruolo che i principi e le definizioni enunciate possono svolgere nel guidare la prassi interpretativa e applicativa, costituendo una base solida per la costruzione di un diritto dell'IA coerente con l'ordinamento costituzionale e multilivello.

Un apposito paragrafo sarà infine dedicato al possibile impatto del Digital Omnibus, il pacchetto legislativo presentato dalla European Commission[4] il 19 novembre 2025 con l'obiettivo dichiarato di razionalizzare e aggiornare la normativa digitale dell'Unione, inclusi alcuni elementi chiave del AI Act e del General

[3] Il principio di precauzione giuridica è stato oggetto di ampie discussioni in dottrina, specialmente per quanto riguarda la sua applicazione in ambito ambientale e tecnologico. Per un contributo ampio sullo stato dell'arte del principio nei vari ambiti del diritto, cfr. Balletti; Foglia (2023).

[4] Le proposte di regolamento dell'Unione europea seguono la procedura legislativa ordinaria. Dopo la presentazione da parte della Commissione europea, il testo viene discusso e modificato dal Parlamento europeo e dal Consiglio dell'UE, che devono raggiungere un accordo comune (tramite negoziati interistituzionali, i cosiddetti "trilogi"). Una volta adottato formalmente da entrambe le istituzioni, il regolamento viene pubblicato nella Gazzetta ufficiale dell'Unione europea ed entra in vigore alla data indicata nel testo stesso, generalmente il ventesimo giorno successivo alla pubblicazione. Essendo direttamente applicabile in tutti gli Stati membri, non richiede recepimento nazionale, salvo eventuali misure interne di adeguamento laddove previste.

Data Protection Regulation (GDPR). Le modifiche prospettate, una volta approvate, inciderebbero direttamente o indirettamente anche sulla Legge n. 132/2025, che è costruita come norma di raccordo e integrazione rispetto al quadro europeo.

2 ARTICOLO 1 – FINALITÀ E AMBITO DI APPLICAZIONE

L'articolo 1 della legge n. 132/2025 svolge una funzione fondativa e programmatica all'interno del quadro normativo italiano sull'intelligenza artificiale. Benché formulato in modo sintetico, esso assume una rilevanza strategica nella costruzione del sistema italiano di governance dell'intelligenza artificiale e va letto come norma di scopo, in senso tecnico: esso orienta l'intero impianto normativo e specifica la ratio sottesa alla legge. In tal senso, assume un ruolo interpretativo e integrativo, in quanto consente di ancorare la disciplina dei singoli articoli (es. art. 3 sui principi generali o art. 6 sugli obblighi dei fornitori) a un disegno politico-legislativo coerente e finalizzato.

Esso definisce, in primo luogo, l'ambito di applicazione della legge, estendendo la disciplina a tutte le fasi del ciclo di vita dei sistemi di IA, dalla ricerca e sviluppo, alla sperimentazione, fino all'adozione e all'impiego operativo. Questa definizione ampia include non soltanto i prodotti finali, ma anche i processi e le attività preliminari, sottolineando l'importanza di un approccio regolatorio che intervenga sin dalle fasi iniziali di progettazione e testing, in coerenza con i principi di "by design" e "by default" sanciti a livello europeo[5].

[5] I principi di "by design" e "by default" sono sanciti dall'art. 25 del Regolamento (UE) 2016/679 (GDPR), che impone agli operatori l'integrazione sin dalla fase di progettazione di misure tecniche e organizzative adeguate per assicurare la protezione dei dati personali e il rispetto della normativa applicabile ("privacy by design"). Inoltre, la protezione deve essere garantita mediante impostazioni predefinite restrittive ("privacy by default"), affinché l'utente non debba intervenire per assicurarsi un livello elevato di tutela. Nell'ambito specifico dell'intelligenza artificiale, tali prin-

Al centro della norma si pone un richiamo esplicito all'uso "corretto, trasparente e responsabile" dell'intelligenza artificiale, inserito in una prospettiva chiaramente antropocentrica. Questo richiamo non rappresenta una semplice asserzione valoriale, bensì svolge una funzione giuridica precisa ovverosia di orientare l'intero impianto normativo alla tutela dei diritti fondamentali della persona e della sua dignità, che trovano tutela nell'articolo 1 della Carta dei diritti fondamentali dell'Unione europea e nella Costituzione italiana, in particolare nei principi espressi negli articoli 2, 3 e 32. In tale prospettiva, la legge italiana si colloca in un filone regolatorio che riconosce l'IA come strumento e mezzo al servizio della persona umana, e non quale entità autonoma o sostitutiva, ribadendo così una concezione etico-giuridica fondata sulla centralità dell'essere umano, e legittimando anche in questo ambito una nozione di antropocentrismo che, nel contesto giuridico, diviene un vero e proprio paradigma normativo che pone la persona umana al centro dell'ordinamento, orientando le norme e le politiche alla protezione e valorizzazione della dignità, dell'autonomia e dei diritti individuali[6].

L'enfasi sull'antropocentrismo si traduce in una serie di obblighi e principi regolatori volti a salvaguardare valori quali la libertà, l'uguaglianza e l'inclusione sociale, prevenendo che lo sviluppo tecnologico possa produrre discriminazioni o violazioni

cipi sono stati estesi e declinati nel Regolamento (UE) 2024/1689 relativo alla disciplina dei sistemi di IA, nonché nelle Linee Guida Etiche sull'Intelligenza Artificiale adottate dall'High-Level Expert Group on AI della Commissione Europea (2019), che sottolineano l'importanza di integrare garanzie di trasparenza, accountability, sicurezza e rispetto dei diritti fondamentali direttamente nella progettazione e nel funzionamento dei sistemi AI.

[6] Sul concetto di antropocentrismo in ambito giuridico e i suoi sviluppi nel diritto tecnologico e dell'IA si vedano, tra gli altri, Floridi (2013), che sottolinea come l'antropocentrismo nel diritto informatico sottenda un orientamento verso la tutela della persona nell'interazione con le tecnologie digitali; Hildebrandt (2015), che approfondisce il rapporto tra tecnologie intelligenti e diritti umani in un'ottica antropocentrica; Sartor (2022) e Sartor; Lagioia (2020) per una riflessione sull'impatto dell'intelligenza artificiale sul diritto e la necessità di un'impostazione normativa che preservi la centralità della persona umana.

dei diritti fondamentali. In questa prospettiva, l'articolo richiama implicitamente la necessità di un bilanciamento tra innovazione tecnologica e tutela dei valori costituzionali e sovranazionali, allineandosi alle più recenti strategie e direttive europee in materia di intelligenza artificiale e tecnologie digitali.

Va inoltre sottolineato che l'approccio antropocentrico e l'importanza attribuita al controllo umano nell'impiego dell'intelligenza artificiale si radicano nel principio giuridico della *riserva di umanità*, che rappresenta un presidio essenziale del nostro ordinamento volto a garantire, nel settore de quo, che le tecnologie non assumano il ruolo di soggetti decisionali autonomi in ambiti che toccano i diritti, le libertà fondamentali e la dignità della persona (Per una compiuta analisi del principio, cfr. Gallone, 2023; Gallone, 2024). La riserva di umanità si configura, in questa prospettiva, non solo come clausola di garanzia, ma come vera e propria condizione strutturale di legittimità dell'agire pubblico e privato assistito da automazione.

La dottrina più recente ha messo in luce come tale principio non possa ridursi a un mero richiamo etico, ma debba essere considerato alla stregua di un criterio giuridico operativo, che impone la presenza di un controllo umano effettivo, e non meramente simbolico o postumo, su tutte le decisioni automatizzate che incidono sulla sfera giuridica individuale. Ciò implica non soltanto la possibilità di intervento, ma anche la concreta attribuzione di responsabilità giuridica a soggetti umani specificamente identificabili per ruolo e posizione. La riserva di umanità rappresenta, dunque, il fondamento di una progettazione regolativa che rifugge modelli di "delegazione cieca" alla macchina, esigendo invece l'integrazione di meccanismi di sorveglianza umana attiva, anche in chiave preventiva[7].

[7] Cfr. Regolamento (UE) 2024/1689, art. 14, par. 4, lett. g).

Il principio della riserva di umanità trova un solido ancoraggio nella Costituzione italiana, in particolare nell'articolo 2, che riconosce i diritti inviolabili della persona, e nell'articolo 3, che vieta ogni forma di discriminazione e richiede l'effettività della tutela (Rodotà, 2012). Ma è soprattutto l'articolo 97, al comma 2, a offrire, sia pure in ottica interpretativa e finalistica, un aggancio diretto per il tema in esame, stabilendo che "l'ordinamento degli uffici, la determinazione delle attribuzioni e la regolamentazione dei ruoli, nonché la definizione delle responsabilità, sono stabiliti per legge o per regolamento". Questa disposizione, tradizionalmente letta in chiave organizzativa, assume oggi un rilievo sistemico alla luce dell'impiego dell'intelligenza artificiale nell'amministrazione pubblica: la chiara determinazione di ruoli e responsabilità, in particolare nei casi di decisioni supportate da IA, diviene infatti condizione necessaria per garantire la trasparenza dell'azione amministrativa, la tracciabilità delle scelte e, soprattutto, la sussistenza di un controllo umano effettivo che impedisca il totale disancoramento del potere decisionale da soggetti giuridici responsabili.

Sotto questo profilo, la riserva di umanità non solo si colloca al crocevia tra organizzazione amministrativa e garanzie procedurali, ma si intreccia con la funzione giurisdizionale, la quale, secondo una larga parte della dottrina (ex coeteribus, Gallone, 2023), non può mai essere automatizzata senza comprometterne l'essenza garantista e il presidio costituzionale dei diritti. Anche nel contesto giurisprudenziale, seppur in via indiretta, si rinviene un orientamento teso a escludere che le decisioni automatizzate possano operare in assenza di verifica, motivazione e revisione da parte di un soggetto umano[8].

[8] Cons. Stato, Sez. VI, 7 dicembre 2022, n. 8472.

Sul piano sovranazionale, il principio di riserva di umanità riceve un'esplicita consacrazione nell'articolo 22 del Regolamento (UE) 2016/679 (GDPR), che vieta le decisioni basate unicamente su trattamenti automatizzati in assenza di tutele adeguate, tra cui il diritto a ottenere l'intervento umano e a contestare la decisione. La tutela contro il rischio di "disumanizzazione" dell'amministrazione digitale viene ulteriormente rafforzata dal nuovo Regolamento (UE) 2024/1689 sull'intelligenza artificiale (AI Act), il quale – nella disciplina dei sistemi ad alto rischio – richiede la presenza di meccanismi di supervisione umana che siano proporzionati, effettivi e idonei a garantire la sicurezza, la trasparenza e la legalità dell'utilizzo dei sistemi. In tale ottica, la legge italiana sull'IA deve essere letta e applicata in chiave sistemica, come parte di un ordinamento multilivello integrato, nel quale le norme nazionali dialogano con i vincoli e le garanzie poste dal diritto europeo.

Ne consegue che ogni decisione pubblica e, per estensione, ogni decisione privata di rilievo giuridico, che si avvalga di strumenti di intelligenza artificiale deve necessariamente contemplare una fase di supervisione umana, idonea a garantirne la legalità, l'efficacia e la compatibilità con i principi costituzionali. La mancata previsione di tale controllo, o la sua mera formalizzazione priva di contenuto sostanziale, determina non solo un deficit procedurale, ma può condurre a vizi di legittimità degli atti amministrativi e, nei casi più gravi, a violazioni dei diritti fondamentali suscettibili di censura costituzionale e sovranazionale.

Per una piena comprensione dell'ambito di applicazione della legge, è inoltre essenziale richiamare la definizione di "sistema di intelligenza artificiale" contenuta nell'articolo 2, su cui si dirà più diffusamente nel prossimo paragrafo, che riprende integralmente quella fornita dall'art. 3, n. 1 del Regolamento (UE) 2024/1689 (AI Act). Tale definizione assume rilievo non solo tecnico, ma

propriamente giuridico, poiché individua l'oggetto specifico della regolazione, delimitando l'ambito soggettivo e oggettivo della disciplina. In base a tale disposizione, un sistema di IA è definito come un sistema progettato per operare con elementi di autonomia e che, ricevendo dati, può inferire in modo automatico obiettivi o output. Si tratta di una formulazione volutamente ampia e flessibile, idonea a ricomprendere una varietà crescente di tecnologie che, anche in assenza di una piena autonomia decisionale, possono produrre effetti significativi su individui, enti e sistemi sociali. In questa prospettiva, il legislatore adotta una nozione funzionale, incentrata sugli effetti dell'output generato dal sistema (decisioni, raccomandazioni, classificazioni, previsioni, ecc.), piuttosto che sulla sua struttura interna o sul tipo di algoritmi impiegati.

Risulta quindi evidente che il campo di applicazione della legge si estende a tutte quelle soluzioni tecnologiche che, operando in modalità automatica o semi-automatica, esercitano un'influenza diretta o indiretta sui comportamenti individuali, sulle decisioni organizzative o sulle dinamiche sociali. Questi sistemi non si limitano a svolgere semplicemente funzioni esecutive o di supporto, ma assumono un ruolo attivo e strategico nella gestione e nell'orientamento di processi complessi, configurandosi come autentici "sistemi di governance tecnologica".

In questa prospettiva, la normativa rompe con la tradizionale visione del software come mero strumento passivo e strumentale, e riconosce invece l'intelligenza artificiale come un soggetto giuridico distinto, che necessita di una disciplina autonoma e specifica. Uno spostamento paradigmatico rivoluzionario che configura l'intelligenza artificiale non più come una semplice estensione tecnologica, ma come un sistema autonomo di governance che richiede un regime giuridico ad hoc, capace di rispondere alle sfide etiche, sociali e giuridiche emergenti dall'adozione su larga scala.

Si comprende bene allora il perché dell'adozione di criteri di responsabilità differenziata, dell'introduzione di obblighi di trasparenza e spiegabilità delle decisioni automatizzate, nonché la previsione di meccanismi di sorveglianza umana e di intervento correttivo nel contesto di una disciplina articolata secondo un approccio basato sul rischio, che consenta di calibrare le misure di prevenzione e controllo in funzione dell'impatto potenziale delle applicazioni AI su diritti fondamentali, sicurezza e principi etici.

L'articolo 1, infatti, nel delineare i principi generali, richiama espressamente la necessità di istituire misure di vigilanza, monitoraggio e mitigazione dei rischi derivanti dall'uso dell'IA, con attenzione particolare ai profili etici, giuridici, sociali ed economici. Questa impostazione riflette l'approccio del diritto dell'Unione Europea, che ha introdotto un modello di regolazione ex ante basato sul rischio (*risk-based regulation*), ispirato a principi di precauzione, proporzionalità e trasparenza[9]. Tale modello si fonda su un principio di precauzione, che impone di agire anche in assenza di certezza scientifica; su un principio di proporzionalità, che commisura gli obblighi normativi al livello di pericolo concreto; e su un principio di trasparenza, che richiede la conoscibilità e comprensibilità dei sistemi da parte degli utenti e delle autorità di controllo.

Il Regolamento (UE) 2024/1689 individua quattro livelli di rischio in relazione ai sistemi di intelligenza artificiale:

- rischio inaccettabile (proibito) proprio di quei sistemi che presentano un impatto gravemente lesivo dei diritti fondamentali e delle libertà civili, come sistemi di manipolazione subliminale, social scoring da parte di autorità pubbliche (sul modello cinese), sistemi di identificazione biometrica remota in tempo reale in spazi pubblici, salvo

[9] Cfr. Considerando n. 30, 34 e 47 del Reg. (UE) 2024/1689.

eccezioni rigorose (esempio pratico: un sistema IA che valuta i cittadini per l'accesso a benefici pubblici sulla base del loro comportamento online o del reddito dei genitori);

- alto rischio. È la categoria centrale del regolamento. Rientrano qui i sistemi impiegati in ambiti sensibili per la persona, come giustizia, occupazione, sanità, educazione, sicurezza, infrastrutture critiche e in generale i sistemi che prendono decisioni automatizzate con effetti legali o simili per le persone. Tali sistemi devono rispettare obblighi rigorosi, tra cui la valutazione della conformità, registrazione, documentazione tecnica, gestione del rischio, audit, sorveglianza umana, ecc.(esempio pratico: un algoritmo usato da un tribunale per suggerire la misura cautelare più adeguata in fase preliminare o un sistema usato da un datore di lavoro per filtrare automaticamente i CV in fase di selezione);
- rischio limitato. In questa fascia rientrano i sistemi che non pongono rischi significativi, ma che possono generare effetti collaterali, ad esempio in termini di trasparenza. Sono richiesti obblighi informativi, come avvisare l'utente che sta interagendo con un sistema di IA o con contenuti generati artificialmente (esempio pratico: una chatbot che risponde a domande del cliente sul sito web di un'azienda o un software di generazione automatica di testi);
- rischio minimo o trascurabile. È la categoria residuale, che copre la maggior parte degli utilizzi quotidiani dell'IA (es. raccomandazioni di contenuti su una piattaforma di streaming, assistenti vocali per uso privato). In questi casi non sono previsti obblighi specifici, ma è incentivata l'adozione volontaria di codici di condotta o pratiche etiche.

Dal punto di vista delle fonti, la legge n. 132/2025 si configura come una norma di principio e di coordinamento, che recepisce e specifica a livello interno gli obblighi derivanti dal diritto dell'Unione, in particolare dal Regolamento (UE) 2024/1689, rispetto al quale riconosce espressamente il primato e il carattere vincolante, ai sensi dell'art. 117, comma 1, Cost. e dell'art. 288 TFUE. In tale ottica, il rinvio alla normativa europea non opera solo come tecnica di richiamo statico, ma come rinvio mobile e dinamico, che impone una lettura sistemica della disciplina e una costante attività interpretativa volta a garantire l'effettività del diritto sovranazionale[10].

In conclusione, l'articolo 1 si configura come una clausola generale di indirizzo che definisce la cornice valoriale, assiologica e regolatoria dell'intera disciplina sull'intelligenza artificiale. Tuttavia, la sua concreta attuazione dipenderà dalla capacità degli operatori – pubblici e privati – e delle autorità di controllo di tradurre questi principi in regole tecniche, protocolli procedurali e presìdi di responsabilità giuridica effettiva. L'efficacia della legge, in ultima analisi, sarà legata alla costruzione di un ecosistema regolatorio interdisciplinare, partecipato e adattivo, capace di tenere insieme innovazione e garanzie, nel pieno rispetto dei diritti fondamentali e della coerenza sistemica dell'ordinamento. In quest'ottica già dalla struttura normativa dell'art. 1 si può enucleare la necessità, nel contesto di un governo giuridico della IA, della costruzione di un sistema di governance multilivello, in cui convergano competenze di autorità nazionali e sovranazionali, enti di normazione per la

[10] Sul tema del rapporto tra diritto interno e diritto comunitario, si veda la storica Corte cost., sent. 5 giugno 1984, n. 170, in Giur. cost., 1984, p. 1613 ss., che ha rappresentato una svolta significativa nel riconoscere la prevalenza del diritto comunitario sulle norme interne incompatibili, pur ribadendo la necessità di un controllo di costituzionalità fondato sull'art. 11 Cost.Più tardi, sul rafforzamento della piena efficacia del diritto comunitario, vedasi Corte cost., sent. 8 giugno 1989, n. 232 che, tra l'altro, chiarisce ulteriormente il dovere del giudice nazionale di disapplicare la norma interna contrastante con il diritto comunitario direttamente efficace.

creazione di standard tecnici (es. CEN-CENELEC- ETSI), organismi di vigilanza e attori economici, secondo una logica di responsabilità condivisa e co-regolazione.

In questo contesto, il principio di legalità richiede una reinterpretazione che tenga conto delle peculiarità e della complessità delle tecnologie emergenti. Non si tratta più di un semplice vincolo formale volto a garantire che le norme siano emanate da un'autorità competente, ma di un principio dinamico che deve assicurare coerenza e adeguatezza tra il quadro normativo e le innovazioni tecnologiche. Pertanto, esso, andando oltre il tradizionale ambito della riserva di legge, finisce con l'estendersi a un sistema integrato e multilivello di regolazione, in cui si combinano regole tecniche, standard europei armonizzati, linee guida e codici di condotta. Questo modello normativo ibrido consente di modulare l'intervento regolatorio in modo flessibile, garantendo al contempo uniformità e certezza del diritto su scala sovranazionale. Gli standard tecnici e le regole armonizzate, in particolare, svolgono un ruolo cruciale nel tradurre i principi giuridici in requisiti operativi concretamente applicabili, facilitando l'adozione di buone pratiche comuni e la certificazione della conformità dei sistemi tecnologici. Le linee guida e i codici di condotta, spesso elaborati da enti di settore o organismi di auto-regolamentazione, arricchiscono tali strumenti fornendo indicazioni più dettagliate e contestualizzate, favorendo così un'effettiva compliance normativa e una governance responsabile.

Si tratta, in definitiva, di un approccio integrato che risponde alla necessità di una regolazione agile e tempestiva, capace di adattarsi rapidamente ai continui sviluppi tecnologici senza rinunciare ai principi fondamentali dello Stato di diritto, quali la trasparenza, la responsabilità e la tutela dei diritti fondamentali. Ne deriva un equilibrio tra certezza giuridica e flessibilità normativa, essenziale per governare l'innovazione in modo sostenibile ed eticamente responsabile.

3 ARTICOLO 2 – DEFINIZIONI

L'articolo 2 della legge n. 132/2025 individua alcune definizioni essenziali per la corretta interpretazione e applicazione della disciplina. Il legislatore nazionale, in una scelta di coerenza sistemica, si limita a richiamare le definizioni contenute nel Regolamento (UE) 2024/1689, riservandosi di integrare soltanto quelle funzionali al contesto interno. La disposizione prevede inoltre una clausola finale di rinvio mobile alle definizioni contenute nel diritto unionale, che si applicano anche per tutti gli aspetti non specificamente disciplinati dalla norma italiana. Le definizioni espressamente richiamate sono le seguenti:

- *sistema di intelligenza artificiale*: "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.";
- *modello di intelligenza artificiale*: " Un modello di IA, anche laddove tale modello di IA sia addestrato con grandi quantità di dati utilizzando l'autosupervisione su larga scala, che sia caratterizzato una generalità significativa e sia in grado di svolgere con competenza un'ampia gamma di compiti distinti, indipendentemente dalle modalità con cui il modello è immesso sul mercato, e che può essere integrato in una varietà di sistemi o applicazioni a valle, ad eccezione dei modelli di IA utilizzati per attività di ricerca, sviluppo o prototipazione prima di essere immessi sul mercato.";
- *dato*: "Qualsiasi rappresentazione digitale di atti, fatti o informazioni e qualsiasi raccolta di tali atti, fatti o

informazioni, anche sotto forma di registrazione sonora, visiva o audiovisiva.”.

La rilevanza giuridica di tali definizioni è duplice; da un lato, esse determinano l’ambito oggettivo di applicazione della normativa, dall’altro, fungono da criterio di qualificazione tecnica e normativa dei soggetti, degli scopi e processi coinvolti.

L’assunzione della definizione di “sistema di IA”, modellata su quella dell’art. 3 AI Act, consente di distinguere nettamente tra strumenti informatici generici e quelli che, per le loro caratteristiche funzionali, sono soggetti a regole di trasparenza, sorveglianza e conformità. La centralità dell’output generato (previsioni, raccomandazioni, decisioni) e la capacità di influenzare l’ambiente operativo (fisico o digitale) rappresentano i tratti distintivi rispetto a software tradizionali.

L’inclusione della definizione di modello di IA rafforza la struttura modulare della disciplina: il modello, in quanto elemento sottostante al sistema, diviene oggetto di attenzione regolatoria autonoma, soprattutto alla luce della crescente diffusione di modelli generativi (es. LLM, foundation models) la cui complessità e opacità sollevano interrogativi in tema di accountability e explainability.

Infine, la definizione di dato riflette un’impostazione ampia e tecnologicamente neutra, coerente con le esigenze del trattamento automatizzato. Ciò è particolarmente rilevante nei contesti in cui la qualità e provenienza dei dati incide direttamente sulla correttezza e imparzialità degli output del sistema IA.

Deve immediatamente osservarsi che l’articolo 2 della legge, pur formalmente limitato alla funzione definitoria, assume una valenza strategica in quanto le definizioni ivi adottate non sono

neutre, ma costituiscono la precondizione tecnica per l'applicazione del sistema di classificazione del rischio. In altri termini, sapere cosa si intende per "sistema di IA", "modello di IA" e "dato" è la base per stabilire a quale livello di rischio esso appartenga e quali obblighi giuridici comporti. La possibilità cioè di attribuire un sistema a un livello di rischio dipende direttamente dalla definizione normativa adottata. Ad esempio, un sistema che usa immagini per il riconoscimento facciale potrebbe essere valutato come ad "alto rischio" solo se rientra nella definizione formale di "sistema di IA" contenuta nel regolamento e, per rimando, nell'articolo 2 della legge italiana. Lo stesso vale per i modelli di IA per finalità generali, che possono assumere livelli di rischio differenziati in base al loro uso effettivo. Inoltre, la definizione nazionale di "dato", molto ampia, influenza il modo in cui si valutano i rischi derivanti dai dataset: se un sistema è addestrato su dati audiovisivi, personali o sensibili, le problematiche di bias, discriminazione, o violazione della privacy devono essere considerate nella valutazione del rischio.

In questo senso, l'articolo 2 non è un semplice articolo tecnico, ma costituisce la base semantico-giuridica della regolazione per livelli di rischio, da esso dipendendo la corretta classificazione e la conseguente applicazione del regime normativo previsto dal regolamento europeo. Il modello risk-based adottato dal diritto europeo e recepito dalla legge italiana impone infatti una forte interconnessione tra principi generali (art. 1), definizioni tecniche (art. 2), e obblighi operativi. È attraverso questa filiera logico-normativa che si può garantire un equilibrio tra innovazione e tutela, tra progresso e controllo, tra efficienza e diritti.

Nel commentare e applicare la legge sull'IA, è dunque essenziale non trattare l'articolo 2 come una clausola neutra o puramente descrittiva, ma come una struttura portante del sistema di classificazione del rischio, la cui esatta articolazione costituisce

la condizione stessa di legittimità e funzionalità della regolazione dell'intelligenza artificiale nel nostro ordinamento.

Sul piano critico, si può osservare che l'adozione di definizioni unionali tramite rinvio comporta un certo grado di instabilità normativa: modifiche future al Regolamento UE o interpretazioni evolutive da parte della Commissione o della Corte di giustizia potrebbero incidere sull'ambito applicativo della legge interna. Inoltre, la genericità di alcune definizioni (in particolare quella di "modello di IA") potrebbe rendere necessaria una futura opera di specificazione settoriale, soprattutto in ambiti ad alta sensibilità (es. sanità, giustizia, lavoro pubblico).

Nel complesso, l'articolo 2 svolge un ruolo cruciale nella costruzione di un lessico normativo condiviso e tecnicamente rigoroso, necessario per governare la complessità e la rapida evoluzione dei sistemi di intelligenza artificiale. La precisione terminologica rappresenta infatti un presupposto indispensabile per una coerenza interpretativa e una corretta applicazione delle norme, elementi fondamentali in un contesto regolatorio così articolato e innovativo.

Tuttavia, si deve rilevare che il ricorso prevalente a definizioni ancorate a fonti europee di carattere generale, come il Regolamento (UE) 2024/1689, espone il sistema normativo a un rischio non trascurabile, quello di non riuscire a cogliere tempestivamente le future innovazioni tecnologiche, soprattutto nel campo tecnologico e dell'IA in continua e rapida trasformazione. In altri termini, la legge potrebbe diventare rapidamente obsoleta o parzialmente inefficace, in quanto le categorie tecniche e i confini definiti oggi potrebbero risultare insufficienti o inadeguati domani.

Questa criticità si inserisce in un fenomeno ben noto alla dottrina giuridica, che lo storico del diritto Giovanni Tamassia

sintetizzava efficacemente affermando che il diritto nasce inevitabilmente "vecchio" rispetto ai fenomeni sociali, economici e tecnologici che intende regolare (Tamassia, 1923). Questa lentezza normativa, intrinseca alla natura stessa del diritto, è particolarmente acuta in settori altamente dinamici come quello digitale, dove la tecnologia evolve spesso a ritmi esponenziali e con modalità imprevedibili.

Per fronteggiare questa sfida, la disciplina dell'intelligenza artificiale richiede una costante opera di revisione, integrazione e specificazione normativa, capace di garantire un delicato e necessario equilibrio tra la flessibilità regolatoria, per adattarsi alle innovazioni e ai cambiamenti tecnologici, e la certezza del diritto, indispensabile per assicurare tutela effettiva dei diritti fondamentali e prevedibilità per operatori e utenti. Solo attraverso un meccanismo dinamico di aggiornamento e governance normativa, che si avvalga anche di contributi interdisciplinari e di un dialogo costante con i settori tecnico-scientifici, sarà possibile mantenere la funzionalità e l'efficacia del quadro regolatorio.

In ultima analisi, la sfida normativa nell'era dell'IA è quella di un diritto proattivo ma prudente, in grado di anticipare senza soffocare l'innovazione tecnologica, di normare senza limitare indebitamente lo sviluppo, e di garantire solidità e adattabilità in un contesto globale caratterizzato da profonde e continue trasformazioni[11].

4 ARTICOLO 3 – PRINCIPI GENERALI

L'articolo 3 della legge italiana sull'intelligenza artificiale (IA), rappresenta il cuore assiologico e normativo della disciplina

[11] Floridi (2013), per cui questa tensione rappresenta la "sfida regolatoria del secolo" che impone di costruire un sistema normativo agile e resiliente, in grado di bilanciare sviluppo tecnologico e tutela dei diritti fondamentali.

nazionale. Esso enuncia sette principi generali che governano l'intero ciclo di vita dei sistemi e modelli di intelligenza artificiale per finalità generali (general-purpose AI), delineando un quadro di garanzie, limiti e finalità costituzionalmente orientato, in armonia con il diritto dell'Unione europea e i trattati internazionali.

4.1 COMMA 1 – I DIRITTI FONDAMENTALI COME ORIZZONTE DELLA TECNICA

Il primo comma dell'articolo 3 stabilisce che la ricerca, la sperimentazione, lo sviluppo e l'utilizzo dei sistemi di intelligenza artificiale devono conformarsi al rispetto dei diritti fondamentali e delle libertà costituzionali, nonché a una serie di principi trasversali: trasparenza, proporzionalità, sicurezza, protezione dei dati personali, riservatezza, accuratezza, non discriminazione, parità di genere e sostenibilità.

Tale disposizione riveste un evidente valore sistemico e finalistico, in quanto individua una vera e propria cornice costituzionalmente orientata allo sviluppo tecnologico, che esclude ogni concezione neutra o meramente tecnocratica dell'innovazione. L'enunciazione dei principi ivi esposta si configura come vincolo giuridico sostanziale, atto a orientare tanto la governance dei sistemi quanto la loro concreta applicazione nei contesti pubblici e privati.

La norma richiama implicitamente un ampio spettro di disposizioni costituzionali: dall'art. 2, che tutela la persona nella sua dimensione inviolabile e relazionale, all'art. 3, in chiave di eguaglianza sostanziale e rimozione delle disuguaglianze, fino agli artt. 13, 15 e 21, che garantiscono la libertà personale, la segretezza delle comunicazioni e la libertà di manifestazione del pensiero. A ciò si aggiungono gli artt. 32 (diritto alla salute), 41 (limite sociale

e ambientale dell'iniziativa economica) e 117 Cost., per il suo riferimento vincolante al rispetto degli obblighi internazionali ed europei, inclusi la Carta dei diritti fondamentali dell'UE e la CEDU.

In questo contesto, il principio di proporzionalità assume un rilievo specifico, configurandosi come criterio-guida per valutare la legittimità dell'impiego dell'IA rispetto agli obiettivi perseguiti. Tale principio, di derivazione euro-unitaria (Palladino, 2024), impone che ogni intervento sia necessario, adeguato e non eccedente rispetto allo scopo, trovando applicazione sia nella dimensione costituzionale interna (art. 97 Cost.) sia nel controllo multilivello dell'azione amministrativa e tecnologica. La proporzionalità opera così come strumento di bilanciamento tra l'innovazione e la salvaguardia dei diritti fondamentali, secondo un modello di "legalità tecnologica" ispirato alla tutela effettiva della persona.

Il richiamo al principio di non discriminazione e parità di genere poggia direttamente sui valori costituzionali e gli obblighi nazionali e sovranazionali che vietano ogni forma di discriminazione, sia essa diretta o indiretta. Tale disposizione assume particolare rilievo nel contesto dell'intelligenza artificiale, settore nel quale si manifestano rischi concreti legati alla presenza di bias algoritmici. Questi ultimi, in presenza di dati di addestramento distorti o parziali, possono tradursi in una riproduzione automatizzata delle disuguaglianze preesistenti, contravvenendo così al principio di uguaglianza sostanziale sancito dall'art. 3 della Costituzione (Peruzzi, 2021). Il principio di uguaglianza sostanziale, infatti, non si limita a vietare discriminazioni formali, ma impone di eliminare gli effetti discriminatori di fatto, un'esigenza che trova una nuova declinazione nelle sfide poste dall'intelligenza artificiale. Il diritto antidiscriminatorio si confronta quindi con la necessità di regolamentare i dati di addestramento e di garantire la trasparenza

e la correttezza dei modelli, al fine di prevenire fenomeni di discriminazione indiretta (Dunn, 2022; Ingrao, 2024, p. 170 e ss.).

Parallelamente, il richiamo alla sostenibilità inserisce nel quadro normativo un'attenzione inedita agli impatti ambientali e sociali dell'intelligenza artificiale.

Sul piano ambientale, le fasi di creazione e training degli algoritmi richiedono un consumo energetico significativo, con conseguenze rilevanti in termini di emissioni di gas serra e impatto ecologico. Ciò impone un'applicazione rigorosa del principio di precauzione e responsabilità, in linea con gli obiettivi di sostenibilità ambientale sanciti da normative europee e internazionali, quali il Green Deal e l'Accordo di Parigi.

Sul piano sociale, l'adozione dell'IA comporta rischi concreti di esclusione digitale, dovuti alla diseguaglianza nell'accesso alle tecnologie e alla disparità nelle competenze digitali. Queste condizioni possono generare nuove forme di marginalizzazione economica e sociale, evidenziando la necessità di un approccio integrato alla sostenibilità che includa anche la dimensione sociale e la governance (Tomasi, 2024, p 47 e ss.).

La dottrina giuridica ha riconosciuto come tali problematiche richiedano un modello di governance responsabile dell'intelligenza artificiale, capace di coniugare principi tradizionali di tutela dei diritti con strumenti di accountability e trasparenza specifici per i sistemi algoritmici (Alvisi; Di Nella, 2024). Questo approccio multidisciplinare è fondamentale per garantire che l'innovazione tecnologica sia coerente con i principi di equità, non discriminazione e sostenibilità.

4.2 COMMA 2 – QUALITÀ, APPROPRIATEZZA E TRASPARENZA DEI DATI

Questo comma si concentra su uno degli assi portanti dell'architettura regolativa dell'intelligenza artificiale: la qualità dei dati. Esso stabilisce che lo sviluppo dei sistemi di IA debba fondarsi su dati e processi che garantiscano correttezza, attendibilità, sicurezza, qualità, appropriatezza e trasparenza, nel rispetto di un principio di proporzionalità settoriale. Tale impostazione implica che la severità dei requisiti debba essere graduata in base all'ambito applicativo e al livello di rischio connesso, secondo una logica coerente con l'approccio risk-based proprio del AI Act europeo.

Il riferimento alla qualità e appropriatezza dei dati comporta, sul piano operativo e giuridico, la necessità di un controllo ex ante sull'intero ciclo della data governance, con particolare attenzione alla provenienza, alla pertinenza, all'affidabilità e alla non distorsività dei dataset. È in questa prospettiva che si inseriscono i concetti di *data lineage* e *data stewardship*, ormai essenziali per garantire tracciabilità, responsabilizzazione e controllo nelle fasi di acquisizione, normalizzazione e conservazione dei dati[12].

In particolare, l'uso di dataset storici o raccolti per finalità divergenti rispetto allo scopo attuale del trattamento espone a rischi

[12] Sul piano tecnico-operativo, i concetti di data lineage e data stewardship costituiscono due elementi centrali della data governance, specie nei settori soggetti a regolamentazione intensiva come l'intelligenza artificiale.Il data lineage consiste nella tracciabilità completa del ciclo di vita del dato, inclusi origine, trasformazioni, trasferimenti e utilizzo, ed è fondamentale per garantire auditabilità, responsabilità e trasparenza nel trattamento algoritmico.La data stewardship, invece, implica l'attribuzione formale di responsabilità per la qualità, la sicurezza e la conformità legale dei dati, assicurando un presidio organizzativo e normativo sul loro corretto utilizzo. Entrambi i concetti sono sviluppati nel contesto degli standard ISO e della normativa europea più recente: ISO/IEC 38505-1:2017 – Information technology – Governance of IT – Governance of data – Part 1: Application of ISO/IEC 38500 to the governance of data, International Organization for Standardization; European Commission, Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance (Data Governance Act);European Commission, High-Level Expert Group on AI – Ethics Guidelines for Trustworthy AI, 2019.

elevati di distorsione predittiva, con possibili effetti discriminatori indiretti. Ciò impone una valutazione non solo tecnica ma anche giuridica dell'appropriatezza funzionale del dato, alla luce del principio di contestualità e minimizzazione già elaborato in ambito privacy dall'art. 5, par. 1, lett. c) del GDPR.

Il richiamo alla trasparenza dei processi non si esaurisce nella pubblicità delle scelte algoritmiche, ma si estende all'intero ciclo di vita del sistema, imponendo che ogni decisione assunta dagli sviluppatori sia adeguatamente documentata, motivata e rendicontabile. Si tratta di un'applicazione evolutiva del principio di accountability, già previsto all'art. 5, par. 2, del GDPR, ora esteso – e rafforzato – nel contesto dell'IA, in particolare per i sistemi ad alto rischio, come previsto dagli articoli 9 (Sistema di gestione dei rischi) e 10 (Dati e governance dei dati) dell'AI Act europeo.

In tale quadro, emerge con chiarezza la necessità di una responsabilità multilivello lungo l'intera filiera dell'intelligenza artificiale, dai fornitori di dataset agli sviluppatori di modelli, fino agli integratori e utilizzatori finali. Il legislatore prefigura così un sistema in cui ciascun attore sia titolare di specifici doveri di diligenza, controllo e rendicontazione, in un regime di responsabilità distribuita, coerente con l'impianto europeo.

Non va infine trascurato il raccordo con i principali standard tecnici e normativi in materia, a partire dagli standard internazionali ISO/IEC (come lo ISO/IEC 25012:2008 sulla qualità dei dati), fino ai recenti Data Governance Act (Reg. UE 2022/868) e Data Act (Reg. UE 2023/2854), che stabiliscono regole comuni sulla condivisione e il riutilizzo dei dati nel contesto dell'economia digitale europea. L'allineamento ai suddetti framework consente non solo una maggiore interoperabilità dei sistemi, ma anche una più robusta tutela degli interessi fondamentali coinvolti.

4.3 COMMA 3 – CENTRALITÀ DELL'UOMO E I PRINCIPI DI CONOSCIBILITÀ, *SPIEGABILITÀ*, SORVEGLIANZA UMANA

Il terzo comma afferma il principio fondamentale della supremazia decisionale dell'essere umano, vincolando la ricerca, lo sviluppo e l'impiego dei sistemi di intelligenza artificiale al rispetto dell'autonomia personale, alla prevenzione del danno e ai requisiti di *conoscibilità*, *spiegabilità* e *sorveglianza umana effettiva*.

Tale previsione normativa risponde sostanzialmente alla problematica nota come *automation bias*, ossia la tendenza, da parte degli operatori umani, a deferire in modo acritico e sistematico le decisioni ai sistemi automatizzati. Tale fenomeno si manifesta quando gli utenti attribuiscono eccessiva fiducia alle raccomandazioni o ai risultati prodotti dall'intelligenza artificiale, riducendo il proprio grado di scrutinio critico e di intervento attivo nel processo decisionale. L'*automation bias* rappresenta un rischio significativo soprattutto in contesti caratterizzati da elevata complessità tecnico-operativa, dove la mole e la sofisticatezza delle informazioni possono sovraccaricare le capacità cognitive degli operatori, portandoli a delegare automaticamente la responsabilità al sistema tecnologico. Inoltre, condizioni di stress, urgenza o pressione decisionale aggravano questo fenomeno, riducendo ulteriormente la capacità di intervento umano consapevole. È il caso, ad esempio, della giustizia predittiva, dove algoritmi complessi suggeriscono decisioni su misure cautelari o valutazioni del rischio di recidiva[13]; dell'ambito

[13] È divenuto famoso al riguardo il caso COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), un sistema automatizzato utilizzato negli Stati Uniti per valutare il rischio di recidiva dei detenuti, influenzando decisioni giudiziarie come la concessione della libertà condizionale e la determinazione delle sentenze. Nonostante l'obiettivo dichiarato di aumentare l'efficienza e l'imparzialità nelle decisioni giudiziarie, numerosi studi e casi giudiziari hanno messo in luce gravi problemi di discriminazione algoritmica, in particolare verso la popolazione afroamericana.Un caso emblematico è rappresentato dall'inchiesta del ProPublica nel 2016, che ha evidenziato come COMPAS tende a sovrastimare il rischio di recidiva per imputati afroamericani, mentre sottostima quello per i bianchi,

medico, in cui sistemi automatizzati possono influenzare diagnosi e terapie; o della sicurezza pubblica, dove decisioni critiche possono essere prese in tempi molto rapidi sulla base di analisi automatizzate di dati sensibili.

L'*automation bias* non solo mina l'autonomia decisionale umana, ma può anche tradursi in conseguenze gravi e ingiustizie sistematiche, come la perpetuazione di errori algoritmici, discriminazioni involontarie e mancate verifiche critiche.Per questo motivo, il legislatore impone che la supremazia decisionale rimanga sempre saldamente nelle mani dell'essere umano, garantendo un'effettiva sorveglianza e un intervento consapevole sulle decisioni automatizzate. Solo in questo modo è possibile mitigare i rischi connessi alla delega automatica e preservare i diritti fondamentali degli individui e a presidio di tale scopo superiore sono posti i requisiti di conoscibilità, spiegabilità e sorveglianza umana.

Il concetto di conoscibilità indica la capacità del sistema di intelligenza artificiale di rendere accessibili e comprensibili le informazioni relative ai processi decisionali automatizzati. Non si tratta soltanto della possibilità tecnica di accedere ai dati o agli algoritmi, ma anche della chiarezza e della trasparenza con cui tali informazioni vengono presentate agli utenti, agli operatori e agli organi di controllo. La conoscibilità si configura quindi come un prerequisito fondamentale per la spiegabilità e per l'esercizio di un controllo efficace sulle decisioni automatizzate, contribuendo a

alimentando così disparità razziali ingiustificate nel trattamento giudiziario. Tali bias sono imputabili a dati storici distorti e a modelli di apprendimento automatico non adeguatamente calibrati per prevenire discriminazioni.La sentenza State v. Loomis (Wisconsin, 2016) ha riconosciuto l'utilizzo di COMPAS come supporto decisionale, ma ha sollevato dubbi sull'opacità del sistema, sulla mancanza di trasparenza nelle sue logiche interne e sulla difficoltà per la difesa di contestare i risultati algoritmici. Questo caso rappresenta un punto di riferimento cruciale nel dibattito giuridico sull'uso dell'intelligenza artificiale nella giustizia, sottolineando l'importanza di garantire spiegabilità, equità e supervisione umana per mitigare i rischi di discriminazione automatizzata.

prevenire fenomeni di opacità algoritmica e favorendo responsabilità, trasparenza e fiducia nell'uso dell'intelligenza artificiale.

Il principio di spiegabilità (explainability) si configura quale garanzia funzionale dell'autodeterminazione informata dell'utente o del soggetto interessato, imponendo che le modalità, i criteri e le logiche sottostanti alle decisioni automatizzate siano rese accessibili, comprensibili e verificabili. Da un punto di vista tecnico, la spiegabilità può essere declinata in vari livelli:

- *spiegabilità tecnica*, rivolta agli sviluppatori e ai regolatori, che riguarda la comprensione degli algoritmi e dei modelli;
- *spiegabilità funzionale*, che si focalizza sulla capacità di illustrare come una determinata decisione viene presa;
- *spiegabilità decisoria*, che rende chiaro all'utente finale il motivo della decisione automatizzata e le sue possibili implicazioni.

Questi livelli sono cruciali per assicurare trasparenza e fiducia nei sistemi di IA, e rappresentano una premessa indispensabile per esercitare un controllo consapevole e responsabile.

Il concetto di "sorveglianza umana" (human oversight), va inteso come strumento di controllo preventivo e correttivo, finalizzato a evitare fenomeni di deresponsabilizzazione e automatismo decisionale incontrollato. La sorveglianza umana deve essere concreta e sostanziale, non meramente formale o simbolica, e deve essere integrata sin dalla fase di progettazione del sistema, in ossequio al principio di human *oversight by design*. Questo approccio implica la predisposizione di meccanismi che consentano all'operatore umano di intervenire efficacemente nelle decisioni automatizzate, correggendo o sospendendo l'azione del sistema quando necessario.

In ambito normativo europeo, il principio della supervisione umana trova espressione nel Regolamento sull'Intelligenza Artificiale (AI Act), che impone l'obbligo di human oversight per i sistemi ad alto rischio, nonché nel Regolamento Generale sulla Protezione dei Dati (GDPR), e in particolare nell'articolo 22, che riconosce il diritto degli interessati a non essere soggetti a decisioni basate unicamente su processi automatizzati, compresa la profilazione, senza un intervento umano significativo.

Questa cornice giuridica sottolinea come la presenza umana lungo tutto l'arco del processo algoritmico non sia un mero requisito formale, ma una condizione sostanziale per garantire la tutela dei diritti fondamentali. Anzi da questo angolo visuale si potrebbe sostenere che all'obbligo di *human oversight* corrisponda un autonomo e nuovo diritto soggettivo digitale alla sorveglianza (Costantino, 2024).

Tale diritto per la sua connaturata attinenza alla salvaguardia della dignità, libertà e sicurezza della persona deve intendersi alla stregua di uno strumento di enforcement dei diritti delle personalità e dei diritti inviolabili della persona nella sfera di rilevanza della relazione digitale, con una disciplina quindi non limitata a quella del Codice Civile o di contingenti normi speciali ma estesa alla garanzia costituzionale per il tramite in particolare dell'art. 2 Cost..

Una lettura che appare ancora più convincente ove si aderisca all'interpretazione di certa dottrina che, superando il tradizionale dibattito sulla natura di norma chiusa o aperta dell'art. 2 Cost., riconosce a quest'ultima il ruolo di vera e primigenia norma tesa a tutelare, non già e non solo i diritti-situazioni giuridiche soggettive attive circoscritti a previsioni specifiche già determinate dal Costituente quanto piuttosto il valore in sé della libera formazione della personalità dell'individuo e pertanto capace di conferire

legittimazione costituzionale anche a quei diritti di nuova emersione e manifestazione nella storia evolutiva dei rapporti sociali e che tuttavia sono impliciti nel o enucleabili dal contesto della Costituzione (Modugno, 1995).

In conclusione, il terzo comma rappresenta una risposta normativa articolata e avanzata alle sfide poste dall'adozione massiva di sistemi di IA, affermando un equilibrio imprescindibile tra innovazione tecnologica e tutela dei diritti fondamentali, attraverso la centralità della supervisione umana, la spiegabilità delle decisioni automatizzate e la prevenzione dei rischi di delega incontrollata alle macchine.

Va però dato conto delle numerose criticità connesse all'effettiva implementazione pratica del principio di *human oversight*. In primo luogo, emerge la necessità imprescindibile di una formazione specialistica e continua per gli operatori umani incaricati di monitorare i sistemi di intelligenza artificiale. Questi devono infatti possedere competenze multidisciplinari che spaziano dalla comprensione tecnica degli algoritmi e dei modelli di machine learning, fino alla conoscenza delle implicazioni etiche e giuridiche delle decisioni automatizzate. La carenza di tali competenze può compromettere la capacità di identificare anomalie, bias o errori nei processi decisionali automatizzati, riducendo l'efficacia del controllo umano.

In secondo luogo, la sorveglianza umana può essere gravata da un significativo sovraccarico cognitivo, soprattutto in contesti caratterizzati da volumi elevati di dati e decisioni da monitorare in tempo reale. Questo sovraccarico può indurre fenomeni di affaticamento mentale e ridurre la vigilanza, con il rischio di trasformare l'intervento umano in una mera formalità piuttosto che in un'effettiva barriera di controllo.

Infine, la letteratura scientifica e giuridica sottolinea i limiti intrinseci dell'intervento umano all'interno di sistemi complessi e altamente automatizzati. L'interazione con algoritmi opachi, spesso descritti come "scatole nere" (black box), rende difficoltoso per gli operatori comprendere appieno il funzionamento interno del sistema e le ragioni precise alla base delle decisioni prese. Ciò può portare a una delega implicita e inconsapevole del potere decisionale, con conseguente deresponsabilizzazione e riduzione della capacità di intervento corretto. Inoltre, l'intervento umano rischia di essere influenzato da bias cognitivi propri, che possono interferire con la capacità di giudizio critico e obiettivo sulle decisioni algoritmiche.

Queste criticità, ampiamente discusse in letteratura (Wachter; Mittelstadt; Floridi, 2017, p. 76) evidenziano come il principio di human oversight debba essere supportato da adeguati strumenti tecnici, organizzativi e formativi, nonché da un quadro normativo chiaro e vincolante che definisca ruoli, responsabilità e modalità di intervento umano nei sistemi automatizzati.

4.4 COMMA 4 – TUTELA DEL METODO DEMOCRATICO E DELLA SOVRANITÀ NAZIONALE

Il quarto comma esplicita una preoccupazione politico-istituzionale connessa all'intrinseca adattabilità dei sistemi di intelligenza artificiale a forme opache di manipolazione, idonee a comprimere in maniera sistematica e massiva le libertà cognitive degli individui. Per tale ragione, il legislatore introduce un principio programmatico fondamentale: l'IA non deve in alcun modo compromettere il metodo democratico, la libertà del dibattito politico né le competenze delle istituzioni territoriali, nel pieno rispetto del principio di sussidiarietà e dell'autonomia degli enti locali.

Tale previsione si inscrive nel solco delle gravi interferenze elettorali e delle manipolazioni cognitive subliminali recentemente denunciate a livello internazionale, grazie all'utilizzo di algoritmi basati su IA nei social network finalizzati a influenzare opinione pubblica e processi elettorali. Non è infatti revocabile in dubbio che specie nell'era algoritmica si assiste sempre più ad una profonda riscrittura delle dinamiche geo-politiche all'interno delle quali le tecniche di comunicazione digitale rivestono un ruolo estremamente critico. Nell'ecosistema comunicativo contemporaneo, infatti, dominato dalle piattaforme digitali e dall'iper-personalizzazione dei contenuti, il profiling e il targeting comportamentale si configurano come strumenti centrali non solo per finalità commerciali, ma anche – e sempre più spesso – per scopi politici, inclusi quelli illeciti o manipolativi.

Il *profiling* consiste nella raccolta e nell'analisi sistematica di dati personali, relativi a comportamenti, interessi, orientamenti ideologici e attività online degli utenti, allo scopo di costruire profili dettagliati e predittivi. Tali informazioni, spesso acquisite mediante piattaforme social o tramite terze parti non sempre trasparenti, permettono di delineare con precisione segmenti di popolazione particolarmente sensibili a determinati temi o narrazioni.

Questo processo di profilazione è propedeutico al *targeting*, ossia alla diffusione selettiva e mirata di contenuti verso individui o gruppi specifici, sulla base delle caratteristiche rilevate. Nel contesto politico, ciò si traduce nella possibilità di indirizzare messaggi personalizzati, spesso sotto forma di "dark ads", ossia annunci visibili solo al destinatario selezionato, che possono avere finalità manipolative, disinformative o polarizzanti.

L'utilizzo congiunto di profiling e targeting apre così la strada a forme di interferenza illecita nel dibattito democratico, poiché

consente a soggetti terzi – talvolta esterni al contesto nazionale – di influenzare clandestinamente l'opinione pubblica e, in ultima istanza, l'esito di processi elettorali. I messaggi veicolati possono essere progettati per rafforzare paure latenti, alimentare divisioni sociali, o minare la fiducia nelle istituzioni democratiche, facendo leva su bias cognitivi e vulnerabilità emotive dell'elettorato.

Un esempio emblematico è rappresentato dal caso Cambridge Analytica[14], in cui dati raccolti in modo opaco sono stati utilizzati per influenzare voti referendari ed elettorali (come nel caso della Brexit o delle presidenziali statunitensi del 2016), dimostrando come l'uso disinvolto di queste tecniche possa alterare l'equilibrio informativo e compromettere la libera formazione dell'opinione pubblica.

Inoltre, il carattere algoritmico e opaco di questi processi rende difficilmente rilevabili tali interferenze da parte delle autorità competenti e degli stessi cittadini. Questo scenario solleva gravi questioni etiche, giuridiche e democratiche, in quanto si assiste a un'inversione del paradigma informativo dal momento che non è più l'elettore a cercare l'informazione, ma è l'informazione, sovente manipolata, a cercare l'elettore, sulla base di un profilo psicografico invisibile e non controllabile. In definitiva, il ricorso ad algoritmi finalizzati a profiling e targeting nelle strategie di comunicazione politica, se non regolamentato da normative rigorose e trasparenti,

[14] Il caso Cambridge Analytica ha rappresentato uno dei più eclatanti scandali legati alla raccolta e all'uso illecito dei dati personali nel contesto politico. Tra il 2016 e il 2018, la società di consulenza politica ha raccolto, senza il consenso esplicito degli utenti, dati di milioni di profili Facebook attraverso un'applicazione terza, sfruttando queste informazioni per realizzare campagne di profilazione psicografica volte a influenzare elettoralmente i cittadini in diversi paesi, tra cui gli Stati Uniti e il Regno Unito. L'operazione ha sollevato preoccupazioni fondamentali su privacy, trasparenza e manipolazione dell'opinione pubblica, evidenziando come le tecnologie digitali possano essere utilizzate per interferire nei processi democratici. Le indagini delle autorità britanniche e statunitensi, insieme ai rapporti giornalistici come quelli del The Guardian, hanno messo in luce pratiche non trasparenti e potenzialmente illecite nell'uso dei dati personali, accelerando il dibattito internazionale sulla regolamentazione della profilazione politica, il targeting elettorale e la tutela dei diritti digitali.

rischia di trasformarsi in un potente strumento di controllo e manipolazione del consenso, ponendo serie minacce alla sovranità del voto e al principio fondamentale di una democrazia informata e partecipata.

Il richiamo dunque al divieto di interferenze illecite, indipendentemente dall'identità del soggetto agente, rafforza la tutela della sfera pubblica democratica contro pratiche manipolative, anche se poste in essere da soggetti privati, delineando un chiaro limite all'uso della profilazione e delle tecniche persuasive automatizzate basate su IA.

4.5 COMMA 5 – COORDINAMENTO CON L'AI ACT

Il quinto comma ha funzione di coordinamento normativo, stabilendo che la legge italiana non introduce nuovi obblighi rispetto al Regolamento (UE) 2024/1689.

Dal punto di vista tecnico, tale previsione ha un effetto limitativo rispetto alla potestà normativa del legislatore nazionale. In sostanza, essa sancisce che l'Italia si limita a garantire l'applicazione del regolamento europeo senza aggiungere, aggravare o ampliare gli obblighi a carico dei soggetti destinatari, coerentemente con la natura direttamente applicabile dei regolamenti UE, che, a differenza delle direttive, non richiedono trasposizione e sono vincolanti in tutti i loro elementi per gli Stati membri.

La funzione del quinto comma è, dunque, di garanzia e di conformità al diritto europeo, limitando ogni eventuale tentazione di introdurre regole sovrapposte o divergenti rispetto al regolamento. In altri termini, agisce come una clausola di rispetto dell'equilibrio normativo europeo, assicurando che la legge nazionale non si trasformi in un veicolo di distorsione dell'impianto regolatorio comune.

Tuttavia,ciò non impedisce al legislatore italiano di esercitare una funzione integrativa e sistematica, specificando modalità attuative o adattamenti organizzativi sul piano interno.

Dal punto di vista processuale, la previsione contenuta nel quinto comma, che esclude l'introduzione di nuovi obblighi da parte della normativa nazionale rispetto a quanto previsto dal Regolamento (UE) 2024/1689, assume una valenza significativa. In particolare, tale clausola incide tanto sull'interpretazione delle norme interne, quanto sull'operatività del principio del primato del diritto dell'Unione nei giudizi dinanzi all'autorità giudiziaria nazionale.

In primo luogo, la clausola costituisce un criterio ermeneutico vincolante per i giudici italiani, che sono chiamati a interpretare le disposizioni della legge nazionale in modo conforme al regolamento europeo. Qualora emergessero dubbi interpretativi circa l'estensione degli obblighi posti dalla legge interna, il quinto comma funge da limite strutturale, imponendo un'interpretazione restrittiva, coerente con il principio per cui i regolamenti europei sono *self-executing* e non devono essere integrati o modificati dal legislatore nazionale.

In secondo luogo, nel caso in cui il legislatore nazionale dovesse violare tale vincolo e adottare norme che introducono obblighi ulteriori o difformi rispetto al regolamento, si aprirebbe la possibilità, in sede processuale, per i soggetti interessati (es. operatori economici, piattaforme digitali, utenti) di chiedere la disapplicazione della norma interna illegittima. Il giudice nazionale, infatti, è tenuto a garantire l'effettività del diritto dell'Unione e, pertanto, deve disapplicare qualsiasi norma interna incompatibile con il regolamento europeo, senza necessità di attendere l'annullamento formale della disposizione.

Tale meccanismo ha importanti ricadute pratiche nei procedimenti giurisdizionali, in quanto consente una tutela immediata

dei diritti garantiti dal regolamento europeo, anche in assenza di un intervento legislativo correttivo. Inoltre, ove sussistano margini di ambiguità o dubbi sulla corretta interpretazione delle disposizioni europee, il giudice nazionale ha la facoltà (o in certi casi l'obbligo) di sollevare una questione pregiudiziale dinanzi alla Corte di Giustizia dell'Unione Europea, ai sensi dell'art. 267 TFUE.

Infine, la violazione del vincolo imposto dal quinto comma potrebbe anche costituire motivo di illegittimità degli atti amministrativi adottati in applicazione di norme interne incompatibili con il regolamento: ciò renderebbe tali atti impugnabili dinanzi al giudice amministrativo per eccesso di potere o violazione di legge, rafforzando la tutela giurisdizionale dei soggetti coinvolti.

4.6 COMMA 6 – CYBERSICUREZZA E RESILIENZA

Il sesto comma introduce un principio di cybersicurezza strutturale, elevandolo a condizione imprescindibile per il rispetto dei diritti e dei principi sanciti dall'articolo 3. Tale disposizione, seppur tecnica, riveste un'importanza cruciale poiché stabilisce che la sicurezza informatica non sia un elemento accessorio, ma un requisito fondamentale lungo tutto il ciclo di vita dei sistemi di intelligenza artificiale. Dal momento della progettazione, passando per l'addestramento e l'utilizzo, fino alla dismissione, le misure di cybersicurezza devono essere proporzionate e basate su un'analisi costante del rischio, adeguandosi dinamicamente alle minacce emergenti.

Questo approccio riconosce la natura complessa e multidimensionale dei rischi associati all'IA. La vulnerabilità di un sistema, infatti, non è mera questione tecnica e può tradursi in violazioni su larga scala dei diritti fondamentali, in distorsione degli output decisionali e persino nell'uso malevolo della tecnologia. Tra i

rischi più noti vi sono gli attacchi adversariali, in cui manipolazioni sottili degli input possono alterare i risultati in modo ingannevole, il *jail-breaking* dei modelli linguistici di grandi dimensioni (Large Language Models), che consente di aggirare le restrizioni di sicurezza e la corruzione dei dati di addestramento, che compromette l'affidabilità e la neutralità dei modelli.

Il sesto comma amplia inoltre il concetto di cybersicurezza ben oltre la mera protezione informatica, includendo la *resilienza*[15] contro tentativi di alterazione e la capacità del sistema di mantenere la propria operatività anche in condizioni avverse. Ciò implica che la sicurezza deve coprire la robustezza logica dei modelli, garantire la continuità operativa e assicurare la possibilità di un auditing permanente. Questi aspetti sono essenziali per tutelare la trasparenza, la tracciabilità e la responsabilità nell'uso dell'IA, elementi fondamentali per la fiducia degli utenti e la correttezza delle applicazioni.

Il principio qui enunciato si inserisce in un contesto normativo europeo ormai consolidato. Regolamenti e direttive come il Regolamento (UE) 2021/694, che istituisce l'Agenzia

[15] La resilienza contro tentativi di alterazione si configura come la capacità intrinseca di un sistema di intelligenza artificiale di resistere, rilevare e mitigare attacchi e manipolazioni esterne che mirano a compromettere l'integrità, la correttezza o la sicurezza dei dati, degli algoritmi e dei processi decisionali automatizzati. Tale resilienza implica l'adozione di misure tecniche e organizzative volte a prevenire o limitare gli effetti di manipolazioni malevole, quali attacchi adversariali, iniezione di dati corrotti o tentativi di compromissione del modello, garantendo così la robustezza e l'affidabilità del sistema. La capacità del sistema di mantenere la propria operatività in condizioni avverse, nota anche come continuità operativa o resilienza funzionale, rappresenta la proprietà del sistema di garantire il regolare svolgimento delle sue funzioni critiche anche in presenza di guasti, attacchi informatici o altri eventi perturbativi. Questa capacità richiede l'implementazione di meccanismi di failover, ridondanza, monitoraggio continuo e risposta automatizzata agli incidenti, assicurando così che le prestazioni e i risultati prodotti dall'IA non subiscano interruzioni o degradazioni rilevanti.In sintesi, tali caratteristiche sono essenziali per assicurare la sicurezza, l'affidabilità e la responsabilità dei sistemi di IA, in particolare quando questi sono impiegati in contesti sensibili o critici dal punto di vista della tutela dei diritti fondamentali e della sicurezza pubblica.

dell'Unione Europea per la Cybersecurity (ENISA), e la recente Direttiva NIS2 (Direttiva (UE) 2022/2555) impongono misure di sicurezza proporzionate e basate sul rischio per tutti gli operatori di infrastrutture critiche digitali, comprese le piattaforme basate su IA. In tale quadro, il sesto comma si allinea alla prassi regolatoria europea, rafforzando la necessità che le piattaforme e i sistemi basati su IA adottino un approccio di risk management continuo, che includa standard minimi quali:

1. la protezione da attacchi adversariali e altre tecniche di manipolazione volte a compromettere l'integrità dei modelli;
2. la salvaguardia della qualità e della sicurezza dei dati di addestramento, evitando la corruzione o la compromissione dei dataset, che possono condurre a decisioni errate o discriminatorie;
3. la garanzia di continuità operativa, affinché eventuali attacchi o malfunzionamenti non si traducano in interruzioni del servizio o perdite di controllo sui processi decisionali automatizzati;
4. l'implementazione di sistemi di monitoraggio e auditing permanente, indispensabili per assicurare la trasparenza e la verificabilità delle scelte operate dall'IA.

In questo senso, il sesto comma non solo recepisce tali standard, ma li applica con specificità al settore dell'intelligenza artificiale, sancendo l'obbligo di una gestione del rischio continua e integrata, configurandosi come un cardine fondamentale per l'armonizzazione tra innovazione tecnologica e tutela giuridica. In conclusione esso sottolinea come la sicurezza digitale non sia solo un requisito tecnico, ma un presupposto essenziale per un utilizzo

etico, affidabile e responsabile dell'intelligenza artificiale, capace di salvaguardare non solo i dati, ma l'intero spettro dei diritti fondamentali nel contesto digitale contemporaneo.

4.7 COMMA 7 – ACCESSIBILITÀ E DISABILITÀ

Infine, il settimo comma garantisce alle persone con disabilità il pieno accesso ai sistemi di intelligenza artificiale e alle relative funzionalità o estensioni, su base di uguaglianza e senza alcuna forma di discriminazione e di pregiudizio, in conformità alle disposizioni della Convenzione delle Nazioni Unite sui diritti delle persone con disabilità, sottoscritta a New York il 13 dicembre 2006, ratificata e resa esecutiva in Italia ai sensi della legge 3 marzo 2009, n. 18.

Questa norma riflette una visione inclusiva dell'innovazione, e richiama l'obbligo positivo dello Stato di rendere le tecnologie accessibili, usabili, compatibili con le esigenze dei cittadini più vulnerabili, anche mediante l'adozione di standard tecnici specifici (es. interfacce vocali, sintesi visive, supporti sensoriali). L'accessibilità, quindi, non è un optional, ma un requisito strutturale e giuridicamente vincolante, che incide sia sulla fase di design che di erogazione dei servizi basati su IA.

5 RILIEVI CONCLUSIVI

Dopo aver approfondito i contenuti e le implicazioni degli articoli 1, 2 e 3, è possibile ricostruire un quadro organico della legge italiana sull'intelligenza artificiale, mettendo in luce come questi articoli si integrino per definire un modello normativo complesso e coerente.

L'articolo 1 della legge italiana sull'intelligenza artificiale si configura come il punto di partenza imprescindibile per comprendere

la visione e la filosofia normativa che informano l'intero testo legislativo. Questo articolo, infatti, non si limita a enunciare principi generali astratti, ma definisce con chiarezza il quadro valoriale in cui deve muoversi lo sviluppo e l'uso dell'IA, ponendo al centro la dignità umana e il rispetto dei diritti fondamentali come linee guida imprescindibili. La legge, dunque, non guarda all'intelligenza artificiale semplicemente come a una tecnologia da disciplinare, ma come a un fenomeno sociale e politico che impatta profondamente sulle modalità di relazione tra cittadini, istituzioni e mercato. L'articolo 1 stabilisce così la priorità della tutela delle persone e delle comunità, sottolineando la necessità di un approccio etico e responsabile che accompagni ogni fase del ciclo di vita delle tecnologie, dalla progettazione all'implementazione, fino alla valutazione degli effetti.

In continuità con questa impostazione valoriale, l'articolo 2 delinea in modo puntuale l'ambito di applicazione della legge e fornisce definizioni fondamentali che costituiscono la base per un regime giuridico chiaro e funzionale. Questo articolo svolge una funzione chiave nell'articolare un sistema normativo capace di adattarsi alla molteplicità e alla complessità delle tecnologie di IA, riconoscendo allo stesso tempo la necessità di salvaguardare i diritti fondamentali quali la privacy, la libertà individuale, la non discriminazione e la sicurezza. La definizione precisa di cosa si intenda per sistemi di IA, nonché l'indicazione dei soggetti obbligati e dei contesti di applicazione, assicurano che la disciplina sia efficace, evitando ambiguità che potrebbero compromettere la tutela giuridica. Inoltre, l'articolo 2 stabilisce un quadro normativo flessibile ma rigoroso, che consente di modulare le misure di prevenzione e controllo in funzione del rischio associato a specifiche applicazioni, in linea con l'approccio basato sul rischio promosso anche a livello europeo.

All'interno della cornice normativa, così impostata dagli articoli 1 e 2, l'articolo 3 rappresenta una disposizione di fondamentale importanza, anzi la più importante, non solo in termini di regolazione tecnica per principi generali, ma anche per il suo valore politico e giuridico. Esso si erge quale elemento reggente di un'architettura normativa che intende coniugare lo sviluppo tecnologico con la salvaguardia dei principi costituzionali, con particolare riferimento alla dignità umana e ai diritti fondamentali. Esso infatti non si limita a stabilire un insieme di principi generali per la gestione dell'intelligenza artificiale, ma sancisce un paradigma di governance che si oppone a qualsiasi forma di approccio tecnocratico o deterministico alla tecnologia, richiamando in modo esplicito il primato della persona e dei diritti umani.

Su un piano di politica legislativa, l'articolo 3 esprime un chiaro orientamento di tutela della libertà, dell'uguaglianza e dei diritti civili, confrontandosi con la sfida di inserire l'IA in un contesto che non comprometta le prerogative democratiche. La norma si inserisce perfettamente nel più ampio contesto normativo europeo, in particolare nel quadro della Strategia Digitale Europea e del Digital Services Act, pur mantenendo una specificità che risponde alla particolare sensibilità della legislazione italiana verso la tutela dei diritti fondamentali e la protezione contro la potenziale invasività della tecnologia. Il principio di proporzionalità, in particolare, che permea l'intero articolo, si configura non solo come un criterio di adeguatezza degli interventi tecnologici, ma come un'efficace protezione contro il rischio che l'evoluzione tecnologica prevalga sulle libertà e sui diritti individuali. L'integrazione dell'IA nella sfera pubblica e privata, secondo i principi delineati nell'articolo, è concepita come un processo che non può prescindere dalla garanzia della trasparenza, della non discriminazione, e del rispetto della dignità umana.

Dal punto di vista assiologico, sistemico e giuridico, l'articolo 3 rappresenta un tentativo di stabilire un equilibrio tra l'efficienza tecnologica e la protezione dei diritti civili, non solo come obiettivo politico, ma come un vincolo giuridico inderogabile. L'articolo 3, a tal fine, in modo esplicito, rinvia a principi ben noti dell'ordinamento costituzionale, come quelli sanciti dagli articoli 2 (tutela della persona), 3 (uguaglianza sostanziale), 13 (libertà personale), 15 (riservatezza delle comunicazioni) e 21 (libertà di espressione) e gli integra con la disciplina dell'IA, rendendo il principio di accountability e la sorveglianza umana le chiavi di lettura per garantire che le scelte tecnologiche siano compatibili con i principi di legalità e giustizia sociale. L'introduzione di misure come la sorveglianza umana attiva e la spiegabilità delle decisioni automatizzate rappresentano strumenti giuridici di controllo che rafforzano la responsabilità e la trasparenza nelle scelte collettive, e non solo la protezione della sfera individuale. Inoltre, il riferimento esplicito al principio di non interferenza illecita nei processi decisionali, con particolare riguardo a contesti elettorali e politici, denota una reazione puntuale alle problematiche sollevate da fenomeni globali di manipolazione algoritmica e influenze esterne tramite tecniche di *profiling* e *targeting*. La legge italiana si pone, così, come argine alle potenziali deviazioni delle tecnologie di IA che potrebbero compromettere l'autonomia dei processi democratici, segnando un confine giuridico preciso contro forme di manipolazione cognitive, da qualsiasi soggetto esse provengano, siano esse pubbliche o private.

Infine, l'articolo 3 si distingue per un approccio complesso e articolato alla regolazione dell'IA, che non mira alla mera disciplina tecnologica ma alla creazione di un quadro normativo che ponga la tecnologia sotto il controllo della legge, orientando responsabilmente il progresso in una direzione compatibile con l'interesse pubblico e il rispetto dei valori democratici con l'obiettivo ultimo di preservare

l'autonomia dell'individuo in un contesto digitale sempre più complesso.

In conclusione, l'articolo 3 non solo risponde alle sfide imposte dalla rapida evoluzione delle tecnologie emergenti, ma riafferma con forza la centralità del diritto e della politica nella definizione e nel controllo delle scelte tecnologiche. Questo approccio legislativo si propone come una delle risposte più avanzate e complete alle sfide globali in ambito tecnologico, configurandosi come un modello normativo che coniuga il progresso tecnologico con la tutela dei valori democratici e dei diritti fondamentali, assicurando che l'innovazione non comprometta mai la giustizia sociale, la sostenibilità giuridica e soprattutto la sovranità digitale dello Stato[16].

Quest'ultima infatti assume un ruolo cruciale nel garantire il controllo democratico sui dati[17], sugli algoritmi e sulle infrastrutture tecnologiche, almeno per due specifiche dimensioni che si integrano a vicenda:

- difendere la capacità delle istituzioni di orientare e regolare le tecnologie nel rispetto degli interessi pubblici e collettivi, preservando la capacità normativa e di governo delle autorità pubbliche, e assicurando la trasparenza, la responsabilità e la tracciabilità dei processi decisionali automatizzati, nel pieno rispetto dei principi di legalità e proporzionalità;

[16] Il concetto di sovranità digitale si riferisce alla capacità degli Stati di esercitare un controllo effettivo e democratico sui dati, sulle infrastrutture informatiche e sugli algoritmi che incidono sulla vita pubblica e privata. La sovranità digitale implica la tutela degli interessi nazionali e collettivi nel contesto della governance tecnologica, garantendo che la regolamentazione delle tecnologie digitali avvenga nel rispetto dei principi costituzionali e dei diritti fondamentali. La tematica è centrale anche nella strategia europea sulla Digital Sovereignty (https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/651992/EPRS_BRI(2020)651992_EN.pdf) e nel quadro normativo del GDPR, che riflette un bilanciamento tra apertura digitale e protezione sovrana dei dati (https://commission.europa.eu/system/files/2020-02/communication-shaping-europes-digital-future feb2020_en_4.pdf).

[17] Floridi (2014) ha approfondito il concetto di "sovranità digitale" e di "etica dell'informazione", sottolineando come il controllo sui dati e sugli algoritmi sia essenziale per la tutela dei diritti in un'epoca digitale.

- attuare modelli concreti di presidio contro le forme di intervento tecnologico opaco e potenzialmente lesivo della pluralità democratica e della libertà individuale, rendendo imprescindibile un sistema di governance multilivello che integri norme nazionali, europee e internazionali, e che consenta un'effettiva partecipazione e controllo democratico sulle innovazioni tecnologiche.

6 GLI IMPATTI POTENZIALI DEL DIGITAL OMNIBUS IN UNA PROSPETTIVA *DE IURE CONDENDO*

La proposta Digital Omnibus della Commissione Europea, che mira a semplificare e riorganizzare una parte significativa del corpus normativo digitale dell'UE, dall'AI Act al GDPR, dalla normativa su cookie e privacy al Data Act fino alle direttive sulla sicurezza digitale, potrebbe incidere direttamente e indirettamente anche sulla legge italiana n. 132/2025. Vediamo quali effetti, in particolare, potrebbero prodursi con riguardo agli articoli 1,2,3.

6.A EFFETTI SUL COORDINAMENTO TRA FONTI EUROPEE E NAZIONALI (ARTICOLO 1)

Uno degli interventi più rilevanti riguarda il rinvio strutturale dell'entrata in vigore delle disposizioni dell'AI Act relative ai sistemi di IA ad alto rischio, collegandone l'effettiva applicazione all'adozione di standard tecnici armonizzati a livello europeo. Il differimento, che posticipa al 2027 il pieno regime applicativo del titolo dedicato ai sistemi ad alto rischio, produce un effetto immediato sul quadro delineato dall'articolo 1 della legge italiana, che si fonda sull'allineamento temporale e funzionale con la disciplina europea.

Il rischio è quello di un disallineamento tra finalità nazionali e tempistiche sovranazionali, con possibili incertezze applicative per gli operatori pubblici e privati chiamati a conformarsi a una normativa il cui presupposto europeo rimane in fase evolutiva.

6.B INCIDENZA SULLE DEFINIZIONI E SUL LESSICO GIURIDICO DELL'IA (ARTICOLO 2)

All'interno del pacchetto Digital Omnibus si ipotizza, nell'ottica di semplificare gli adempimenti per le imprese e favorire l'innovazione, una revisione del perimetro definitorio del dato personale, soprattutto in relazione ai criteri di identificabilità e ai dati pseudonimizzati i quali ultimi potrebbero eventualmente essere esclusi dall'ambito di tutela qualora non sussista un rischio concreto di re-identificazione. Un cambiamento che, se confermato, avrebbe ripercussioni rilevanti sulla disciplina italiana dell'IA, poiché la Legge 132/2025 si fonda sull'interoperabilità concettuale con il GDPR e con l'AI Act, e quindi su definizioni e presupposti derivati dagli standard europei, e presuppone un livello elevato di protezione dei dati personali e sensibili. Una definizione più restrittiva potrebbe infatti indebolire garanzie essenziali come trasparenza, informazione, consenso e condizioni di trattamento, riducendo l'effettività dei principi generali della legge relativi alla dignità e ai diritti fondamentali.

Parallelamente, il pacchetto rafforza la distinzione tra categorie di sistemi di IA e modelli di base ai fini della responsabilità e della conformità. Poiché l'articolo 2 della Legge n. 132/2025 riproduce e integra la terminologia dell'AI Act, una modifica del vocabolario normativo europeo impone una riflessione sulla tenuta delle definizioni nazionali. Una ridefinizione europea più flessibile del concetto di dato personale o delle categorie di rischio potrebbe

produrre un effetto a cascata sul sistema definitorio italiano, ponendo il legislatore nazionale di fronte alla necessità di aggiornare le proprie categorie o, alternativamente, di mantenere una disciplina più rigorosa, con possibili frizioni interpretative.

Esse potrebbero risultare ridondanti, incomplete o perfino in contrasto con la nuova prospettiva europea, esponendo il legislatore italiano all'obbligo di successivi interventi correttivi o interpretativi.

6.C IMPATTO SUI PRINCIPI GENERALI NAZIONALI (ARTICOLO 3)

Al tempo stesso, il pacchetto introduce semplificazioni e alleggerimenti per imprese e PMI attraverso meccanismi di conformità proporzionata, riduzioni degli obblighi documentali e l'istituzione di sandbox regolatori. Queste misure, pur potendo favorire l'adozione dell'IA da parte di operatori più piccoli, rischiano di entrare in tensione con i requisiti di accountability, trasparenza, non discriminazione, e tutela dei diritti posti dall'articolo 3 della Legge n. 132/2025, che li considera valori-guida.

Se tali semplificazioni dovessero tradursi in minori obblighi di rendicontazione e verifica, la capacità della legge di fungere da presidio di garanzia potrebbe risultarne compromessa o indebolita. Tutto ciò incide anche sul principio di precauzione giuridica che costituisce la struttura portante della legge italiana. Un impianto costruito per evolversi gradualmente attraverso decreti attuativi, regolamenti secondari, prassi e giurisprudenza rischia infatti di andare "fuori sincrono" rispetto a un quadro sovranazionale che cambia rapidamente. La conseguenza potrebbe essere la necessità di riconsiderare norme secondarie, attuazioni e interpretazioni, mettendo in discussione la stabilità normativa che la legge intendeva

assicurare. Non a caso, diverse associazioni per i diritti digitali avvertono che il Digital Omnibus potrebbe rappresentare un vero e proprio "rollback" delle protezioni: la combinazione di ritardi, ridefinizioni di dato personale e semplificazioni operative rischia di rendere l'AI Act, e, per riflesso, le normative nazionali di raccordo, meno vincolante e più permissivo, con conseguente riduzione dell'efficacia di tutela nei contesti ad alto impatto come salute, lavoro, credito e servizi pubblici, proprio quelli che la legge 132/2025 intendeva salvaguardare attraverso un approccio antropocentrico e responsabile.

In questo scenario, per l'Italia si potrebbero delineare due percorsi possibili: da un lato, una revisione o un adeguamento della legge n. 132/2025 per mantenere coerenza con il nuovo quadro europeo, mediante modifiche legislative, aggiustamenti dei decreti attuativi o misure transitorie; dall'altro, il rischio di un disallineamento, in cui la legge resterebbe formalmente in vigore ma con garanzie concrete indebolite, poiché gli operatori potrebbero sfruttare le deroghe e le semplificazioni introdotte a livello UE.

In entrambi i casi emerge chiaramente che la normativa nazionale non può essere interpretata come un contenitore chiuso, ma come parte di un ecosistema multilivello condizionato dalle evoluzioni europee, il che richiede vigilanza continua, dialogo tra istituzioni, autorità, operatori e società civile e, forse, un ripensamento dei meccanismi di governance dell'IA nel nostro ordinamento.

In conclusione, il Digital Omnibus rappresenta un banco di prova cruciale per la tenuta del modello italiano di governance dell'intelligenza artificiale e pone la Legge 132/2025 davanti a un crocevia, potendo trasformarla in uno strumento destinato ad adattarsi a un contesto europeo più permissivo, con conseguente perdita di efficacia, oppure offrire l'occasione per consolidare un approccio

nazionale più coerente e robusto. L'impianto basato su precauzione e principi generali potrebbe così evolvere verso una governance attiva e dinamica, rafforzata attraverso un uso consapevole degli strumenti multilivello di armonizzazione e capace di reagire alle evoluzioni europee e, allo stesso tempo, di preservare attraverso una scelta politica consapevole i valori fondativi della disciplina italiana dell'intelligenza artificiale.

RIFERIMENTI

ALVISI, Chiara; DI NELLA, Luca. **Intelligenza artificiale sostenibile e diritto civile**. Napoli: ESI, 2024.

BALLETTI, Emilio; FOGLIA, Laura (a cura di). **Le dimensioni giuridiche del principio di precauzione**. Napoli: ESI, 2023.

COSTANTINO, Fortunato. Nuovi diritti soggettivi digitali: il diritto di disconnessione e il diritto al controllo umano sull'algoritmo. un tentativo di ricostruzione sistematica in una prospettiva costituzionalmente orientata. In: **Diritto di Famiglia e delle Persone**, Milano, Ed. Giuffrè, anno LIII, Fasc. 4, 2024.

DUNN, Pietro. Moderazione automatizzata e discriminazione algoritmica: il caso dell'hate speech. In: **Rivista Italiana di Informatica e Diritto**, Firenze, Ed. Istituto di Informatica Giuridica e Sistemi Giudiziari del CNR, n. 1, 2022.

FLORIDI, Luciano. **The ethics of information**. Oxford: Oxford University Press, 2013.

FLORIDI, Luciano. **The fourth revolution**: how the infosphere is reshaping human reality. Oxford: Oxford University Press, 2014.

GALLONE, Giovanni. **Riserva di umanità e funzioni amministrative: indagine sui limiti dell'automazione tra**

procedimento e processo. Padova: Cedam, 2023.

GALLONE, Giovanni. Riserva di umanità, intelligenza artificiale e funzione giurisdizionale alla luce dell'IA Act. In: **Judicium**, Pisa, Ed. Pacini, 2024.

HILDEBRANDT, Mireille. **Smart technologies and the end(s) of law**. Cheltenham: Edward Elgar, 2015.

INGRAO, Alessandra. Critica della ragione artificiale: la discriminazione algoritmica intersezionale. In: **Rivista Giuridica del Lavoro e della Previdenza Sociale**, Roma, Ed. Tecnindustria, n. 2, 2024.

MODUGNO, Franco. **I "nuovi diritti" nella giurisprudenza costituzionale**. Torino: Giappichelli, 1995.

PALLADINO, Rossana. **Il principio di proporzionalità nel diritto dell'Unione europea**: natura, funzioni e controllo. Bari: Cacucci, 2024.

PERUZZI, Marco. Il diritto antidiscriminatorio al test di intelligenza artificiale. In: **Labour & Law Issues**, Bologna, Ed. Alma Diamond, Alma Mater Studiorum Università di Bologna, v. 7, n. 1, 2021.

RODOTÀ, Stefano. **Il diritto di avere diritti**. Bari / Roma: Laterza, 2012.

SARTOR, Giovanni. **L'intelligenza artificiale e il diritto**. Torino: Giappichelli, 2022.

SARTOR, Giovanni; LAGIOIA, Francesca. Le decisioni algoritmiche tra etica e diritto. In: RUFFOLO, Ugo (Ed.). **Intelligenza artificiale**: il diritto, i diritti, l'etica. Milano: Giuffrè, 2020.

TAMASSIA, Nino. **Storia del diritto italiano**: storia delle fonti dall'eta romana ai tempi nostri, Padova: La Litotipo, 1923.

TOMASI, Marta. Intelligenza artificiale, sostenibilità e responsabilità intergenerazionali: nuove sfide per il costituzionalismo? In: **Rivista AIC**, Roma, Ed. Associazione Italiana dei Costituzionalisti, n. 4, 2024.

WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. In: **International Data Privacy Law**, Oxford, Ed. Oxford University Press, p. 76-99, 2017.

Recebido em: 22-12-2025

Aprovado em: 22-2-2026