

L 132/2025



ANALISI L. 132/2025

Quali sono le criticità dal punto di vista di chi gestisce la materia tecnica, buchi della norma, ambiti dove la norma non copre i rischi e consigli tecnici.



NICOLINI MASSIMILIANO

FONDAZIONE OLIVETTI TECNOLOGIA E RICERCA - OLITEC

Ottobre 2025

Sommario

PREMESSA GENERALE.....	3
CAPITOLO I – PRINCIPI E FINALITÀ (Artt. 1–3)	8
Centralità antropocentrica e trasparenza algoritmica	8
Proporzionalità e contesto d’uso	9
Sorveglianza umana e autonomia computazionale	10
Cybersicurezza come condizione pregiudiziale	11
CAPITOLO II – DATI, PRIVACY E TRASPARENZA (Artt. 4–6).....	15
Ridefinizione del concetto di dato	15
Pseudonimizzazione e re-identificazione	17
Consenso informato e complessità algoritmica	20
CAPITOLO III – SANITÀ, RICERCA E DISABILITÀ (Artt. 7–10).....	22
L’IA come co-decisore clinico	22
Validazione scientifica dei modelli	25
Dati sanitari e anonimizzazione.....	28
CAPITOLO IV – LAVORO, DIRITTI E GIUSTIZIA (Artt. 11–15)	31
L’IA nel lavoro	31
Giustizia e automazione	34
CAPITOLO V – DATI, ALGORITMI E METODI MATEMATICI (Artt. 16–17)	37
CAPITOLO VI – CYBERSICUREZZA E DIFESA (Art. 18)	39

CAPITOLO VII – STRATEGIA NAZIONALE E AUTORITÀ (Artt. 19–20)	43
Governance e istituzioni	43
Sovranità tecnologica	46
CAPITOLO VIII – SETTORI APPLICATIVI E PROMOZIONE (Artt. 21–23) ..	49
CAPITOLO IX – DELEGHE, FORMAZIONE E FUTURO DEL LAVORO (Art. 24)	52
CAPITOLO X – CONCLUSIONI GENERALI	55

Trattato di Analisi e Interpretazione Tecnica della Legge n.132/2025

L'intelligenza artificiale tra diritto, tecnologia e sovranità umana

3

PREMESSA GENERALE

La **Legge 23 settembre 2025, n. 132**, può essere considerata un testo di svolta nella cultura giuridica italiana, perché segna il primo tentativo organico di costruire una cornice normativa per la **ricerca, lo sviluppo e l'applicazione dell'intelligenza artificiale**. Essa non nasce nel vuoto, ma nel pieno di una fase di profonda ridefinizione geopolitica e tecnologica, in cui il potere delle grandi piattaforme globali — statunitensi e asiatiche in primis — ha iniziato a produrre un effetto di dipendenza cognitiva e infrastrutturale sull'intero continente europeo. L'Italia, con questa legge, tenta di reagire a tale dinamica proponendo un modello che integra la visione comunitaria dell'**AI Act** (Regolamento UE 2024/1689) con una serie di principi nazionali che riflettono la tradizione umanistica e costituzionale del Paese: la tutela della persona, la salvaguardia della dignità, la libertà del pensiero e la responsabilità etica dell'innovazione.

È una norma che nasce dunque in un **momento di vulnerabilità strategica**. Il rischio che l'Europa — e in particolare l'Italia — diventi un semplice utilizzatore di tecnologie progettate altrove non è più ipotesi teorica, ma scenario concreto. I grandi modelli linguistici, le reti neurali generative, le piattaforme di analisi predittiva sono quasi interamente controllate da soggetti extracomunitari che detengono sia la proprietà intellettuale dei sistemi, sia il controllo delle infrastrutture di calcolo. In

tale contesto, la legge italiana si pone un obiettivo ambizioso: **conciliare la crescita tecnologica con la difesa della sovranità cognitiva**, costruendo un equilibrio tra la spinta innovativa e la protezione dei diritti fondamentali.

4

La legge 132/2025 si articola come **una legge di principi**, non di strumenti. Essa definisce un orizzonte etico e politico, ma non ancora una serie di meccanismi tecnici in grado di governare la complessità della nuova intelligenza artificiale. È un testo che parla di valori — centralità umana, trasparenza, sicurezza, sostenibilità, equità — ma che non traduce ancora questi concetti in protocolli operativi. Da un lato, sancisce l'importanza della **formazione algoritmica**, della trasparenza dei dati e della cooperazione pubblico-privato per la cybersicurezza; dall'altro, manca di una struttura di **verifica, di audit e di validazione tecnica** dei modelli che ne garantisca l'applicazione reale.

Questa distanza tra principio e prassi è il cuore della tensione che attraversa l'intero impianto legislativo. Il diritto, per sua natura prescrittivo e lineare, fatica a contenere un oggetto che è **adattivo, probabilistico e autogenerativo**. I sistemi di IA contemporanei — e in particolare i modelli generativi multimodali — non si limitano a eseguire istruzioni, ma apprendono, combinano e producono conoscenza nuova. Si tratta di entità epistemiche che evolvono in modo autonomo, spesso al di fuori della possibilità di controllo umano diretto. In questo scenario, le categorie giuridiche classiche — responsabilità, causalità, imputabilità — entrano in crisi, perché non riescono più a catturare la natura dinamica e non lineare dei processi di apprendimento automatico.

Le aree in cui la **tecnologia già supera la legge** sono numerose e decisive.

La **generatività dei modelli** ha infranto il paradigma della prevedibilità normativa: una macchina capace di scrivere, disegnare, decidere e negoziare non è più un semplice strumento, ma un soggetto cognitivo distribuito.

L'**autonomia decisionale** dei sistemi di IA, utilizzati in settori come la finanza, la sanità e la giustizia predittiva, pone interrogativi radicali sulla natura del giudizio e sul principio di responsabilità. La **manipolazione dei dati sintetici**, ormai indistinguibili dai dati reali, mette in crisi il concetto stesso di prova e di veridicità informativa.

Infine, la **velocità evolutiva delle architetture di apprendimento** — che cambiano e si auto-modificano più rapidamente di quanto una legge possa essere aggiornata — evidenzia la necessità di un diritto dinamico, capace di adattarsi, non di inseguire.

Per queste ragioni, la Legge 132/2025 necessita di un **secondo livello attuativo**, una strategia strutturale che traduca i principi in meccanismi permanenti di vigilanza, adattamento e verifica. È indispensabile costruire una **Strategia Nazionale di Sorveglianza Algoritmica**, un sistema pubblico di controllo e monitoraggio continuo dei modelli impiegati, in grado di garantire trasparenza, affidabilità e accountability. Tale strategia dovrebbe basarsi su tre pilastri:

– un regime di **audit continuo**, che consenta di testare e certificare i modelli nel tempo, verificando l'aderenza alle norme e la sicurezza cognitiva dei sistemi;

- una politica di **trasparenza open-source**, che renda pubblici gli elementi fondamentali dei modelli impiegati nella pubblica amministrazione e nelle infrastrutture critiche;

- la creazione di un **registro nazionale dei modelli algoritmici**, contenente dati su architetture, dataset di addestramento, livelli di rischio e metriche di affidabilità.

Solo un'infrastruttura del genere può trasformare la sorveglianza da principio etico in **pratica democratica**, capace di proteggere i cittadini senza soffocare l'innovazione. L'obiettivo non deve essere il controllo autoritario della tecnologia, ma la sua **integrazione responsabile** nella società, in un equilibrio tra efficienza e trasparenza, tra progresso e garanzia dei diritti.

Il futuro dell'intelligenza artificiale, dunque, non potrà essere ridotto a una questione di controllo giuridico: sarà un problema di **convivenza cognitiva**. Il diritto dovrà imparare a dialogare con sistemi che non obbediscono più alla logica binaria, ma a quella **probabilistica e adattiva**. Le leggi del futuro non potranno più essere statiche, dovranno diventare **leggi in apprendimento**, capaci di evolvere parallelamente alle tecnologie che regolano.

In questa prospettiva, la Legge 132/2025 rappresenta un inizio, non un punto d'arrivo. È il segnale che l'Italia ha compreso la necessità di collocare la persona al centro dell'era digitale, ma è anche la prova che il vero terreno della sovranità non è più solo giuridico, bensì **infrastrutturale e cognitivo**. Solo quando il diritto saprà convivere con l'intelligenza artificiale come con un interlocutore, e non come con un oggetto, potrà nascere una **nuova giustizia algoritmica**, fondata non

sulla paura della macchina, ma sulla consapevolezza che la macchina, per quanto intelligente, resta una proiezione dell'uomo — e che l'uomo, se formato, libero e consapevole, resta il suo custode.



CAPITOLO I – PRINCIPI E FINALITÀ (Artt. 1–3)

Centralità antropocentrica e trasparenza algoritmica

La legge dichiara che l'IA deve essere sviluppata in una dimensione antropocentrica, cioè centrata sull'uomo, ponendo la persona e la sua dignità come fine ultimo dell'innovazione. Tuttavia, questa visione si scontra con la natura stessa dei moderni sistemi di deep learning, i quali non operano in termini semantici o concettuali ma elaborano correlazioni numeriche ad alta dimensionalità, prive di un senso ontologico direttamente interpretabile. Le reti neurali profonde costruiscono rappresentazioni statistiche del mondo, non categorie logiche, e il significato che producono è un effetto emergente piuttosto che un'intenzione. Parlare di "spiegabilità" diventa dunque un esercizio più filosofico che tecnico, poiché il funzionamento interno di un modello di linguaggio di grandi dimensioni (LLM) o di una rete neurale convoluzionale è frutto di miliardi di pesi interconnessi che generano comportamenti non linearmente prevedibili. Anche gli stessi autori dei modelli possono al massimo descriverne la struttura e il processo di addestramento, ma non spiegare in modo esaustivo perché un modello scelga una determinata risposta piuttosto che un'altra. L'antropocentrismo giuridico si confronta quindi con una macchina post-simbolica che sfugge ai criteri tradizionali di causalità e comprensione, imponendo la necessità di un nuovo linguaggio interpretativo per la relazione uomo-algoritmo. In questo contesto, la cosiddetta "spiegabilità" dovrebbe essere ripensata non come capacità di tradurre in linguaggio umano i processi interni di un modello, ma come costruzione di strumenti epistemologici intermedi, capaci di rappresentare la logica statistica della macchina attraverso mappe

concettuali, gradienti di inferenza o visualizzazioni cognitive. La legge, se volesse davvero affrontare il tema, dovrebbe distinguere tra trasparenza tecnica (documentazione del codice e del training), spiegabilità epistemica (capacità di comprendere le relazioni causa-effetto), e comprensibilità sociale (possibilità per l'utente di attribuire significato e fiducia all'esito). L'attuale formulazione si limita a un ideale etico privo di metodo. In realtà, la spiegabilità è un continuum che va dalla matematica alla percezione: ciò che oggi si definisce "black box" potrebbe, con opportuni strumenti, diventare una "gray box" parzialmente illuminata. La ricerca contemporanea sta infatti esplorando approcci di interpretabilità controfattuale, gradient-based e simbolici, ma nessuno di essi garantisce una trasparenza totale. Per questo motivo, la spiegabilità dovrebbe essere concepita come un diritto procedurale, non come una condizione ontologica, richiedendo che ogni sistema espliciti i propri limiti di conoscenza, gli ambiti di validità e le zone d'incertezza, restituendo così al cittadino non l'illusione della comprensione, ma la consapevolezza critica del funzionamento algoritmico.

Proporzionalità e contesto d'uso

Il principio di proporzionalità mira a bilanciare innovazione e sicurezza, ma presuppone la conoscenza anticipata dei rischi. In realtà, il rischio algoritmico emerge *ex post*, cioè dopo che il sistema ha interagito con dati reali e situazioni impreviste, evolvendo in modo dinamico e spesso non lineare. Questo rende impossibile una valutazione preventiva completa: l'IA, specialmente nelle sue forme generative e adattive, può alterare le proprie traiettorie di inferenza in risposta a nuovi input o condizioni di contesto. Ciò significa che i rischi non sono

semplicemente prevedibili ma emergenti, e si sviluppano come proprietà del sistema nel tempo.

10

Un quadro di proporzionalità efficace dovrebbe dunque prevedere non solo una valutazione *ex ante*, ma un regime di sorveglianza continua, capace di monitorare in tempo reale il comportamento dei modelli, analizzare deviazioni e attivare misure correttive automatiche. La legge, invece, resta ancorata a una visione statica del controllo, basata su autorizzazioni e verifiche periodiche, senza prevedere meccanismi adattivi di controllo continuo, come sistemi di *risk tracking*, monitoraggio post-deployment o revisione dinamica dei parametri. Sarebbero indispensabili protocolli di audit algoritmico permanente, dotati di infrastrutture di calcolo nazionali, in grado di verificare la coerenza tra principio e realtà e di reagire rapidamente ai mutamenti del rischio. Solo un approccio dinamico di questo tipo può garantire una vera proporzionalità nell'era dei sistemi evolutivi.

Sorveglianza umana e autonomia computazionale

L'articolo 3 prevede che i sistemi di IA operino sotto sorveglianza e intervento umano. Tuttavia, nelle architetture distribuite e nei modelli auto-addestranti, il controllo umano è solo nominale. L'algoritmo può modificare il proprio comportamento in base a nuovi input, generando decisioni non tracciabili. La sorveglianza umana, dunque, diventa simbolica, non effettiva. In molti casi, l'intervento umano si limita a una supervisione superficiale o a un controllo documentale dei risultati prodotti dal sistema, senza un reale potere di incidere sulla logica interna del processo algoritmico. Nelle architetture di IA distribuite e nei modelli auto-addestranti, l'autonomia computazionale diventa un elemento strutturale, poiché le macchine sono progettate per

apprendere e adattarsi in modo indipendente all'ambiente operativo. Ciò determina una progressiva delega cognitiva dall'uomo alla macchina: l'essere umano non governa più la decisione, ma si limita ad accettarne l'esito.

11

Questa evoluzione apre un conflitto ontologico e giuridico. Da un lato, il diritto presume che ogni atto decisionale abbia un soggetto responsabile e identificabile; dall'altro, la tecnologia produce decisioni emergenti da un sistema collettivo di agenti non pienamente prevedibili. È qui che il concetto di 'sorveglianza' va ridefinito: non come presenza occasionale dell'uomo nel ciclo decisionale, ma come integrazione continua di consapevolezza umana nel flusso cognitivo della macchina. La futura normativa dovrà prevedere meccanismi di controllo adattivo, interfacce di responsabilità condivisa e protocolli di arresto o revisione automatica in presenza di comportamenti anomali.

Senza tali misure, la sorveglianza si riduce a una forma rituale, un atto di fede tecnologica che maschera la realtà di sistemi auto-organizzati. In una prospettiva più ampia, la sfida non è soltanto tecnica ma epistemica: occorre educare operatori, medici, giuristi e ingegneri a comprendere il funzionamento profondo dell'autonomia computazionale, creando una cultura della vigilanza intelligente. Solo così si potrà preservare il principio costituzionale di responsabilità personale anche nell'era dell'azione algoritmica autonoma.

Cybersicurezza come condizione pregiudiziale

La legge riconosce correttamente che la cybersicurezza rappresenta il prerequisito fondamentale della fiducia nei sistemi di intelligenza artificiale, ma nel farlo conserva una visione statica, quasi immobile, di

un concetto che per sua natura è invece dinamico e in continua evoluzione. La sicurezza di un sistema algoritmico non è una condizione permanente, ma un equilibrio fragile che muta con il tempo, con l'ambiente informatico e con le tecniche d'attacco che si perfezionano ogni giorno. Pensare la cybersicurezza come un requisito verificato una volta per tutte equivale a ignorare la realtà della tecnologia contemporanea, in cui le vulnerabilità possono emergere improvvisamente, anche in modelli apparentemente stabili e certificati.

Un sistema di intelligenza artificiale può essere sicuro oggi e diventare vulnerabile domani. Gli attacchi informatici si sono evoluti al punto da penetrare non più soltanto i server o le reti, ma gli stessi processi di apprendimento del modello. È il caso del data poisoning, in cui un aggressore inserisce in modo subdolo dati manipolati all'interno dei dataset di addestramento, modificando gradualmente il comportamento del modello fino a renderlo inattendibile o, nei casi più gravi, malevolo. Si tratta di una forma di corruzione silenziosa che può rimanere latente per mesi e manifestarsi solo in determinati contesti operativi, quando il sistema comincia a fornire risposte anomale o a violare principi di equità e imparzialità. Parallelamente, esistono le tecniche di inversione del modello, mediante le quali un attaccante riesce a risalire ai dati sensibili utilizzati per l'addestramento. In questo modo, il segreto industriale e la privacy degli individui vengono compromessi, mostrando quanto labile sia il confine tra sicurezza informatica e tutela dei diritti fondamentali.

Di fronte a questa realtà, la legge dovrebbe prevedere meccanismi di audit continuo, capaci di verificare nel tempo la sicurezza e l'affidabilità dei sistemi di IA. Gli attuali protocolli normativi si concentrano quasi

esclusivamente sulla fase di pre-deployment, cioè prima che il sistema venga messo in funzione, ma non considerano la necessità di un controllo costante e ciclico durante tutto il ciclo di vita operativo. Un modello di audit dinamico dovrebbe monitorare costantemente l'integrità dei dati, le modalità di accesso, i comportamenti anomali del sistema e la coerenza delle sue risposte rispetto ai parametri originari di progetto. La sicurezza algoritmica non può esistere senza un processo di revisione continua e senza un ecosistema normativo che riconosca la necessità di aggiornare periodicamente le certificazioni di conformità.

Un altro limite della legge risiede nell'assenza di sandbox pubbliche, ambienti controllati e aperti alla ricerca nei quali sia possibile sperimentare e testare i modelli di IA in condizioni simulate, senza rischiare la compromissione dei dati o la violazione delle normative. Le sandbox permetterebbero a università, enti di ricerca e autorità indipendenti di verificare la robustezza dei modelli, valutando come reagiscono a input imprevisti o a tentativi di manipolazione. Queste piattaforme costituirebbero una vera e propria palestra etica e tecnica, in grado di accrescere la trasparenza e la fiducia collettiva nella tecnologia, poiché la fiducia non si costruisce con le dichiarazioni ma con la possibilità di verificare.

Una normativa moderna dovrebbe quindi superare la logica del controllo a tantum e introdurre un sistema di certificazioni dinamiche, con obblighi di aggiornamento periodico basati sul livello di rischio del modello e sul contesto in cui esso opera. Dovrebbe promuovere reti di cooperazione pubblico-privata per la condivisione delle vulnerabilità emergenti e favorire la creazione di laboratori di sicurezza algoritmica che fungano da garanti indipendenti, così come oggi avviene per i

laboratori di analisi farmaceutica o per i centri di omologazione dei dispositivi medici. Solo un tale ecosistema permetterebbe di coniugare l'innovazione tecnologica con la salvaguardia della fiducia pubblica.

14

In definitiva, la legge si limita a riconoscere l'importanza della cybersicurezza come principio, ma non ne istituzionalizza gli strumenti operativi necessari per garantirla nel tempo. Rimane così ferma alla dimensione della fiducia dichiarata, priva di quella verifica costante che sola può trasformarla in fiducia dimostrata. La sicurezza, nel mondo dell'intelligenza artificiale, non è una conquista ma un processo: richiede vigilanza, trasparenza e una responsabilità condivisa tra produttori, enti di controllo e utenti finali. Solo in questa prospettiva la fiducia potrà evolvere da semplice condizione di adesione formale a valore reale, misurabile e continuamente rinnovato.

CAPITOLO II – DATI, PRIVACY E TRASPARENZA (Artt. 4–6)

15

Ridefinizione del concetto di dato

La formulazione della legge, ancorata a una concezione tradizionale del “dato” come informazione esplicitamente identificabile (ad esempio nome, indirizzo, codice fiscale, immagine, suono o testo riconducibile a una persona), mostra oggi tutta la sua inadeguatezza rispetto alla complessità del dato computazionale contemporaneo. Nel dominio dell’intelligenza artificiale, infatti, l’unità di informazione non è più rappresentata soltanto da un valore osservabile o archiviato, bensì da **entità matematiche derivate** – *embeddings*, *feature vectors*, rappresentazioni latenti – che traducono aspetti del mondo reale in spazi multidimensionali di numeri. Queste entità, pur non contenendo direttamente il “nome” o la “faccia” di una persona, racchiudono relazioni, correlazioni e tratti distintivi sufficienti a **ricostruire o inferire identità, comportamenti e stati psicofisici** con un grado di accuratezza sorprendente.

L’aspetto più critico risiede nel fatto che tali rappresentazioni non rientrano formalmente nella categoria di *dato personale*, poiché non sono immediatamente interpretabili dall’uomo. Tuttavia, da un punto di vista tecnico, esse sono **proiezioni funzionali del dato originario**: lo contengono, lo trasformano, ma non lo annullano. Un *embedding* ottenuto da una fotografia, ad esempio, può essere sufficiente per ricostruire il volto originale o per identificare un soggetto in un dataset diverso tramite algoritmi di similarità. Un *feature vector* proveniente da dati sanitari può consentire a un modello di riconoscere predisposizioni genetiche o patologie, anche se l’elemento originario (il nome o la cartella clinica) è stato rimosso.

Qui si apre una contraddizione profonda tra **diritto e matematica**: la legge riconosce e tutela ciò che è esplicitamente nominabile, ma il calcolo statistico e l'apprendimento automatico operano nel dominio dell'implicito, del dedotto, del rappresentato. In altre parole, la **trasformazione del dato non ne annulla la natura personale, ma ne nasconde la tracciabilità dietro una maschera di astrazione**. Questa opacità concettuale è oggi il principale punto cieco delle normative in materia di protezione dei dati: si protegge ciò che è visibile, mentre il rischio si nasconde nei livelli intermedi di rappresentazione algoritmica.

L'assenza di una definizione giuridica di *dato derivato* o *dato rappresentazionale* comporta diverse criticità operative:

Impossibilità di audit completo dei modelli, perché i vettori di rappresentazione non vengono considerati dati sensibili e dunque non soggetti a obblighi di tracciamento, crittografia o cancellazione su richiesta.

Vulnerabilità alla reidentificazione, poiché gli *embeddings* possono essere incrociati tra modelli diversi, permettendo di risalire a individui anche in presenza di pseudonimizzazione.

Incoerenza nella catena di responsabilità, dato che il produttore del modello può sostenere di non trattare dati personali, mentre di fatto manipola rappresentazioni derivate che mantengono legami semantici con le persone da cui provengono.

Inadeguatezza delle misure di consenso, poiché l'utente concede l'autorizzazione per l'uso dei propri dati, ma non per le infinite

rappresentazioni che ne derivano nel tempo attraverso addestramenti successivi e riutilizzi incrociati.

17

In sintesi, la legge tutela un concetto di dato **statico e identificabile**, mentre la realtà computazionale è **dinamica e inferenziale**: ciò che conta oggi non è più il dato in sé, ma la capacità di generare conoscenza da strutture che lo rappresentano in modo astratto. Senza un aggiornamento normativo che includa anche le rappresentazioni derivate — con meccanismi di certificazione, accountability e diritto alla “cancellazione semantica” — si rischia di proteggere un archivio vuoto, mentre il reale contenuto informativo continua a vivere e a propagarsi nei modelli, invisibile ma potenzialmente dannoso.

In conclusione, la distanza tra **dato giuridico e dato computazionale** rappresenta una delle più gravi fratture del diritto digitale contemporaneo: il legislatore resta fermo alla nozione di archivio, mentre l'intelligenza artificiale ha già spostato la frontiera verso spazi di rappresentazione in cui ogni informazione diventa una traccia numerica della nostra identità cognitiva, comportamentale ed emotiva. Colmare questa distanza non è solo una necessità tecnica, ma un dovere etico per garantire la **dignità e la libertà informativa dell'essere umano nell'era degli algoritmi**.

Pseudonimizzazione e re-identificazione

La pseudonimizzazione nasce come uno strumento concepito per separare il dato dall'identità, nella convinzione che la rimozione dei riferimenti diretti come nome, codice fiscale o indirizzo potesse rendere il contenuto informativo innocuo dal punto di vista personale. Si tratta di

un principio figlio di un'epoca in cui la minaccia era considerata statica e il pericolo derivava da un accesso non autorizzato ai dati originari. Ma l'evoluzione dell'intelligenza artificiale ha completamente sovvertito questa logica, perché oggi la re-identificazione non dipende più dal possesso del dato sorgente, bensì dalla capacità di un modello di ricostruire, inferire e collegare frammenti di informazione apparentemente neutri.

Le intelligenze artificiali moderne riescono a ricomporre identità anche partendo da elementi parziali, come una sequenza di movimenti, una voce sintetizzata, un tracciato cardiaco o uno stile linguistico. I sistemi di analisi multimodale riescono a fondere dati provenienti da domini diversi — immagini, testi, sensori, cronologie di comportamento — restituendo un profilo individuale altamente preciso anche quando le informazioni di partenza risultano pseudonimizzate. Il processo di correlazione, reso possibile dal *transfer learning* e dal calcolo statistico ad alta dimensionalità, consente di associare con rapidità crescente ciò che il legislatore presumeva separato. Di conseguenza, la pseudonimizzazione non è più una barriera reale, ma un filtro fragile, trasparente per le capacità inferenziali dei modelli.

L'aspetto più critico è che la legge continua a considerare questo meccanismo una misura di sicurezza sufficiente, accettando che i dati pseudonimizzati possano essere usati per la ricerca o l'innovazione senza consenso esplicito, purché non sia "probabile" risalire all'identità di una persona. Questa idea di probabilità, valida nel paradigma pre-algoritmico, è diventata inadeguata: oggi le correlazioni che un modello può generare trasformano la probabilità in certezza. La distinzione fra dato anonimo e dato pseudonimo perde così significato, poiché una

macchina può ricostruire connessioni latenti con livelli di accuratezza che sfuggono al controllo umano.

19

La conseguenza diretta è una profonda disarticolazione del sistema di tutele. Il concetto di anonimato perde stabilità tecnica: nessun dato può dirsi davvero anonimo, perché ogni informazione, se combinata con altre, può diventare un identificatore. L'apparente neutralità di un dataset è solo una condizione transitoria. La responsabilità giuridica rimane affidata al metodo anziché al risultato, come se l'applicazione formale della pseudonimizzazione bastasse a garantire la sicurezza. In realtà il rischio non risiede più nel metodo, ma nella capacità delle macchine di superarlo. A questo si aggiunge una crescente asimmetria di potere informativo: enti pubblici o privati dotati di modelli sofisticati possono ricostruire identità, comportamenti e condizioni sanitarie anche da dati ritenuti "neutri", aggirando ogni principio di consenso.

La legge, ferma su schemi statici, ignora questa nuova forma di intelligenza emergente, una capacità di re-identificazione autonoma che rende obsoleto il concetto stesso di anonimizzazione. La protezione della privacy non può più essere affidata a tecniche di mascheramento applicate a posteriori, ma deve essere incorporata nei modelli, nei loro algoritmi e nei processi di apprendimento. La privacy deve diventare una proprietà del sistema, non un accessorio giuridico. Solo un'evoluzione normativa che riconosca la natura inferenziale del dato e che consideri il rischio come un prodotto della capacità cognitiva delle macchine potrà restituire equilibrio tra tutela individuale e progresso scientifico. Oggi, ciò che appare protetto lo è solo per l'occhio umano, non per l'intelligenza artificiale che, silenziosamente, continua a ricostruire, riconoscere e classificare.

Consenso informato e complessità algoritmica

Il consenso informato, così come definito nelle normative sulla protezione dei dati, si fonda sull'idea che l'utente possa esercitare una scelta libera e consapevole riguardo al trattamento delle proprie informazioni, dopo averne compreso le modalità, le finalità e i rischi. Questo presupposto, tuttavia, entra in crisi nel momento in cui il trattamento non è più eseguito da sistemi deterministici o da operatori umani, ma da modelli di intelligenza artificiale caratterizzati da un numero immenso di parametri e da una logica di apprendimento non lineare. In un contesto in cui il comportamento dell'algoritmo è il risultato di processi probabilistici, retroazioni interne e dinamiche di ottimizzazione non esplicitabili, la **comprensione reale** da parte dell'utente è semplicemente impossibile.

L'asimmetria cognitiva che si crea tra il cittadino e il sistema algoritmico è assoluta: da una parte vi è un individuo che riceve un'informazione semplificata, dall'altra un apparato che elabora correlazioni su scala planetaria, utilizzando modelli linguistici o predittivi con miliardi di parametri interni e reti di dipendenze che nessun essere umano, nemmeno i suoi creatori, è in grado di spiegare completamente. Questa distanza non è solo tecnica, ma epistemologica: l'utente non può comprendere non perché sia ignorante, ma perché **il linguaggio dell'algoritmo è intrinsecamente opaco**, fondato su rappresentazioni matematiche che non hanno un corrispettivo diretto nel linguaggio naturale.

In questo scenario, la trasparenza diventa una forma di cortesia formale, non una garanzia sostanziale. Le informative di consenso, anche quando dettagliate, descrivono l'intenzione del trattamento, non il suo funzionamento reale. Dire a un utente che un sistema “analizza i suoi dati per migliorare l'esperienza d'uso” non equivale a spiegargli che il modello può inferire il suo stato emotivo, ricostruire le sue abitudini di consumo, anticipare decisioni o dedurre condizioni di salute a partire da correlazioni statistiche che nemmeno gli sviluppatori riescono a decifrare. L'idea stessa di consenso informato si svuota di senso quando l'informazione fornita è incompleta per natura, perché ciò che l'algoritmo apprende non è mai interamente conoscibile.

Si tratta di una trasformazione profonda del rapporto tra individuo e tecnologia: l'utente non partecipa più come soggetto informato, ma come **fonte di dati all'interno di un ecosistema cognitivo che lo supera**. La scelta che compie — accettare o rifiutare — si fonda su una fiducia implicita nel sistema, non su una conoscenza reale. In questo modo il consenso non è più un atto di volontà consapevole, ma una forma di **adesione passiva**, una delega in bianco fondata sulla speranza che l'uso dei propri dati non venga distorto.

Questa condizione produce un effetto paradossale: la trasparenza, anziché tutelare, diventa un rituale burocratico. I documenti di policy, le informative e le clausole di accettazione vengono ridotti a formule che trasferiscono la responsabilità dal titolare al cittadino, il quale però non possiede né gli strumenti né la capacità cognitiva per valutare il rischio. La libertà di scelta è sostituita da un'apparente partecipazione, e la tutela del consenso si trasforma in una **forma di deresponsabilizzazione istituzionalizzata**.

In tale contesto, continuare a invocare la trasparenza come principio cardine del diritto digitale equivale a mantenere un’illusione di controllo in un ambiente che è strutturalmente incontrollabile. Nessuna forma di informativa può tradurre l’opacità di un modello generativo o predittivo in conoscenza realmente accessibile all’utente. La vera sfida, oggi, non è tanto rendere l’algoritmo “comprensibile”, quanto ridefinire il significato stesso di consenso nell’era dell’opacità algoritmica, restituendo al cittadino un potere decisionale effettivo che non dipenda dalla capacità di capire ciò che nessuno, nemmeno i progettisti, può più spiegare in modo esaustivo.

CAPITOLO III – SANITÀ, RICERCA E DISABILITÀ (Artt. 7–10)

L’IA come co-decisore clinico

La legge, nel suo impianto attuale, ribadisce con forza un principio di fondo: la decisione medica deve restare in capo all’uomo, quale ultimo garante della vita e della dignità del paziente. Tuttavia, questa affermazione, apparentemente rassicurante, si scontra con la realtà operativa della medicina contemporanea, dove i sistemi di intelligenza artificiale non fungono più da meri strumenti di supporto, ma da veri e propri co-decisorie nel processo diagnostico. In radiologia, cardiologia e oncologia, gli algoritmi di analisi delle immagini e i modelli predittivi intervengono non solo nell’assistenza interpretativa, ma nella formulazione stessa dell’ipotesi clinica, modificando i percorsi decisionali del medico senza che questi ne abbia piena consapevolezza.

L'errore di prospettiva della normativa è quello di considerare l'intelligenza artificiale come un utensile, un'estensione tecnica dell'occhio o della mano del medico, quando in realtà si tratta di un **soggetto cognitivo complementare**, capace di generare inferenze, proporre priorità diagnostiche e, di fatto, orientare la scelta terapeutica. In molti reparti ospedalieri, il referto iniziale viene prodotto dal sistema di IA e solo successivamente validato dal medico, che spesso si limita a confermare l'esito più probabile indicato dalla macchina. L'atto medico, dunque, non è più un atto solitario, ma un processo di **co-decisione epistemica**, dove l'uomo interpreta e filtra una proposta già formulata da un'intelligenza autonoma.

In questa dinamica, la distinzione giuridica tra “supporto” e “decisione” perde significato. Il potere dell'algoritmo non risiede nell'emettere un verdetto finale, ma nell'orientare la percezione di verità del medico, alterando il peso cognitivo delle alternative. Quando un sistema diagnostico segnala, con un indice di probabilità superiore al 95%, la presenza di una lesione neoplastica, il professionista, anche se formalmente libero, è psicologicamente e statisticamente vincolato da quella evidenza. Di fatto, **l'IA impone un contesto epistemico che condiziona la decisione**, introducendo un vincolo invisibile di fiducia tecnica.

Da questo mutamento emerge una frattura normativa rilevante. La legge continua a imputare la responsabilità all'uomo, ignorando che la decisione finale è il risultato di un'interazione sinergica in cui la macchina ha contribuito in modo determinante alla costruzione del giudizio clinico. La responsabilità esclusiva del medico non rispecchia più la realtà dei processi sanitari contemporanei, dove ogni diagnosi è il

prodotto di un ecosistema computazionale fatto di reti neurali, dati storici e modelli adattivi che evolvono nel tempo. Ciò genera un vuoto giuridico profondo: chi risponde se l'algoritmo sbaglia? Il medico che si è fidato? Il produttore del software? L'ospedale che ne ha autorizzato l'uso? Oppure il modello stesso, come entità operativa che ha contribuito all'atto decisionale?

La soluzione non può essere la mera estensione della responsabilità medica all'uso delle tecnologie, poiché questo equivarrebbe a **negare la natura cognitiva dell'intelligenza artificiale**. Serve un salto concettuale: la creazione di una nuova categoria giuridica che riconosca la *responsabilità condivisa uomo-macchina*. In essa, il giudizio medico non viene annullato, ma contestualizzato all'interno di una relazione di cooperazione epistemica. La macchina non è un agente morale, ma è un agente causale: produce effetti, genera inferenze, condiziona scelte. Ignorarlo significa lasciare senza risposta una porzione sempre più ampia dell'attività clinica.

Una simile categoria dovrebbe riconoscere la **dualità dell'atto medico contemporaneo**, che non è più solo empirico ma anche computazionale. La responsabilità condivisa implicherebbe un sistema di tracciabilità delle decisioni, una documentazione continua delle interazioni uomo-macchina e un regime assicurativo capace di ripartire il rischio tra gli attori coinvolti. Allo stesso tempo, imporrebbe agli sviluppatori un obbligo di trasparenza algoritmica proporzionato alla criticità del contesto clinico, e agli ospedali un dovere di monitoraggio etico e tecnico delle prestazioni dei sistemi impiegati.

Questo nuovo quadro permetterebbe di superare la finzione giuridica che oggi separa l'atto umano da quello algoritmico. Nel momento in cui una decisione medica nasce dall'interazione tra competenze biologiche e intelligenze artificiali, non si può più parlare di autonomia unilaterale, ma di **co-autorialità decisionale**. L'etica e il diritto devono dunque evolvere verso un modello di corresponsabilità, in cui l'uomo rimanga garante del senso e del fine dell'atto clinico, ma la macchina venga riconosciuta come parte attiva della costruzione del sapere medico. Solo in questo modo la legge potrà ritrovare aderenza alla realtà scientifica e proteggere davvero, e non solo formalmente, la dignità della cura nell'era dell'intelligenza artificiale.

Validazione scientifica dei modelli

La previsione normativa che impone verifiche periodiche dei modelli di intelligenza artificiale impiegati in ambito sanitario rappresenta, almeno in apparenza, una misura di garanzia. Tuttavia, dietro questa formulazione si nasconde una profonda fragilità strutturale: la legge enuncia l'obbligo ma non ne definisce né la metrica né la cornice operativa. Non vengono stabiliti criteri di valutazione, non viene indicata la frequenza dei controlli e, soprattutto, non viene identificata con chiarezza l'autorità responsabile della supervisione. Questo vuoto normativo produce una zona grigia dove la conformità formale sostituisce la verifica sostanziale, e dove la qualità e l'affidabilità dei modelli rischiano di diventare elementi aleatori, variabili da struttura a struttura, da software a software.

L'assenza di standard comuni si traduce in una disomogeneità pericolosa all'interno dei sistemi sanitari nazionali. Ogni ospedale, ogni regione, ogni fornitore può interpretare in modo diverso cosa significhi "verifica periodica": per alcuni potrebbe essere un controllo annuale sul funzionamento generale del software, per altri un audit interno sui parametri statistici di performance, per altri ancora una semplice dichiarazione di conformità del produttore. In mancanza di protocolli uniformi, la comparabilità dei risultati e la coerenza diagnostica tra strutture diverse diventano impossibili. Ciò significa che due pazienti, con la stessa patologia e la stessa tipologia di esame, potrebbero ricevere diagnosi differenti non per una differenza clinica, ma per una differenza regolatoria o tecnica nella gestione dei modelli.

Il problema è aggravato dalla natura dinamica dell'intelligenza artificiale. Un modello che apprende nel tempo, che si aggiorna con nuovi dati o che adatta i propri pesi in base all'esperienza, non può essere trattato come un dispositivo statico. La sua affidabilità varia con l'evoluzione del contesto e del dataset, e ogni aggiornamento può introdurre bias, variazioni di accuratezza o instabilità nei risultati. Senza un quadro di riferimento preciso che disciplini la periodicità e la metodologia delle verifiche, diventa impossibile garantire la stabilità epistemica del sistema diagnostico. La legge chiede un controllo, ma non specifica *che cosa* debba essere controllato: l'accuratezza del modello, la sua robustezza agli attacchi, la trasparenza dei pesi, l'equità dei risultati o la coerenza con i protocolli clinici? L'indeterminatezza semantica del termine "verifica" lascia spazio a una molteplicità di interpretazioni, aprendo la porta a pratiche arbitrarie e a rischi non gestiti.

A questo si aggiunge un ulteriore punto di criticità: la mancanza di un'autorità competente centralizzata. In assenza di un ente che definisca criteri tecnici, parametri di accettabilità e procedure di audit, la responsabilità della verifica ricade di fatto sugli operatori sanitari o sulle singole strutture, che non dispongono delle competenze necessarie per valutare modelli di apprendimento profondo da milioni o miliardi di parametri. Ciò genera una responsabilità indiretta ma gravosa per i medici, chiamati a fidarsi di strumenti che la normativa li obbliga a verificare ma non li abilita a comprendere. È una contraddizione strutturale che mina il principio stesso di tutela del paziente, trasferendo il rischio da chi progetta il sistema a chi lo utilizza.

Questa lacuna produce effetti a catena anche sul piano giuridico. In caso di errore diagnostico, diventa impossibile attribuire con chiarezza la colpa: se il modello non è stato aggiornato correttamente, è responsabile il produttore o la struttura sanitaria che non ha condotto le verifiche? E se la legge non indica né modalità né cadenza di tali controlli, come si può imputare una violazione oggettiva? L'assenza di standard condivisi genera un vuoto di accountability che rischia di paralizzare le procedure legali e assicurative. In mancanza di una cornice verificabile, l'errore diventa un evento senza responsabile certo, un'anomalia che si dissolve nella complessità tecnica del sistema.

Il risultato complessivo è una perdita di coerenza e di fiducia. I sistemi sanitari, invece di diventare più sicuri grazie all'intelligenza artificiale, rischiano di frammentarsi, producendo differenze di trattamento e diseguaglianze diagnostiche. La legge, nel tentativo di normare l'innovazione, si limita a una prescrizione formale che, priva di strumenti di attuazione, diventa una dichiarazione di principio senza potere

operativo. La coerenza scientifica e giuridica dei sistemi sanitari richiede un linguaggio normativo capace di tradurre in parametri oggettivi ciò che oggi resta sospeso tra burocrazia e algoritmi. Solo allora la “verifica periodica” potrà trasformarsi da formula retorica a garanzia effettiva della sicurezza del paziente e della responsabilità del medico.

Dati sanitari e anonimizzazione

L'articolo 8, nel prevedere la possibilità di utilizzare dati anonimi per finalità secondarie — di ricerca, di sviluppo o di statistica — parte da un presupposto che appartiene a un'altra epoca della tecnologia: che l'anonimato possa essere garantito in modo stabile e definitivo. Questo principio, apparentemente razionale, si dissolve non appena viene confrontato con la realtà operativa delle reti neurali profonde e dei sistemi di apprendimento automatico. L'idea di un dato “anonimo”, privo di qualunque legame con la persona da cui proviene, non regge più nel contesto computazionale contemporaneo. Le reti neurali non si limitano a memorizzare, ma **apprendono strutture statistiche e pattern fisiologici unici**, traducendo caratteristiche corporee, vocali o comportamentali in rappresentazioni numeriche che, pur prive di etichette identificative, possono condurre all'individuo con un grado di precisione sempre crescente.

Quando un modello di intelligenza artificiale viene addestrato su immagini mediche, tracciati cardiaci, registrazioni vocali o movimenti oculari, esso interiorizza regolarità che non sono solo cliniche, ma biometriche. Queste regolarità costituiscono una sorta di *impronta digitale algoritmica*, capace di distinguere un soggetto da un altro anche

in assenza di metadati personali. L'anonimizzazione tradizionale — basata sulla rimozione dei nomi, dei codici e dei riferimenti diretti — non cancella questa impronta, perché il corpo stesso, con le sue proporzioni, i suoi ritmi, le sue variazioni fisiologiche, è già un identificatore. In altre parole, **la persona è iscritta nei dati** in modo indelebile, e la macchina lo sa riconoscere anche se l'uomo non può più farlo.

La legge continua tuttavia a distinguere rigidamente tra dato personale e dato anonimo, affidando a questa separazione giuridica la legittimità dell'uso secondario. È qui che si apre il paradosso più profondo: ciò che viene dichiarato "anonimo" secondo il linguaggio legale non lo è più secondo quello matematico. L'anonimizzazione diventa un mito giuridico, una finzione che permette di superare la barriera etica del consenso individuale. Si afferma che il dato non è più riconducibile a un soggetto, e dunque può essere liberamente riutilizzato; ma in realtà il dato conserva dentro di sé un'impronta unica, latente, pronta a riemergere non appena venga incrociata con altre fonti o elaborata da modelli più potenti.

La fragilità dell'anonimato si manifesta in modo drammatico proprio nei contesti in cui la precisione dell'analisi è più alta. Un sistema di *deep learning* addestrato su scansioni cerebrali può riconoscere lo stesso individuo attraverso esami differenti, separati da anni e privi di etichette, semplicemente perché il cervello possiede un pattern anatomico personale. Allo stesso modo, una rete neurale che analizza la voce può identificare chi parla anche dopo l'applicazione di filtri di anonimizzazione. L'anonimato, lungi dall'essere una condizione

assoluta, diventa così **una proprietà relativa, revocabile e temporanea**, che cessa di esistere non appena la tecnologia avanza.

30

Questa condizione genera una contraddizione insanabile nel diritto dei dati. L'articolo 8 concede l'uso libero dei dati anonimi perché presume l'assenza di rischio per l'individuo. Ma se l'anonymato non è tecnicamente garantibile, l'uso secondario diventa un trattamento potenzialmente invasivo, mascherato da neutralità. L'assenza di identità nominale non equivale all'assenza di identità computazionale. Ogni modello, ogni rete, ogni archivio digitale è in grado di ricostruire una forma di riconoscimento che sfugge alla logica giuridica, ma che permane nel dominio numerico dei vettori e delle correlazioni.

In questa prospettiva, l'anonimizzazione non è più una protezione, ma un linguaggio di legittimazione. Serve a dichiarare conformità, non a garantire sicurezza. È il modo con cui il diritto tenta di mantenere il controllo su una realtà che gli è sfuggita di mano. Ma la verità è che **non esistono più dati veramente anonimi**, perché ogni traccia digitale porta con sé la possibilità della re-identificazione. Anche quando le informazioni vengono decontestualizzate, rimane una firma algoritmica che le collega al soggetto originario.

La conseguenza è che l'intero impianto normativo che si fonda sulla distinzione tra dato personale e dato anonimo risulta compromesso. Continuare a parlare di "uso secondario dei dati anonimi" significa perpetuare un concetto giuridico che non trova più riscontro nella realtà scientifica. Finché la legge non riconoscerà questa trasformazione, continuerà a proteggere un'illusione, mentre la conoscenza automatica — discreta, silenziosa, ma inesorabile — continuerà a penetrare gli spazi

più intimi dell'identità umana, ricostruendola anche là dove il diritto dichiara che è stata cancellata.

CAPITOLO IV – LAVORO, DIRITTI E GIUSTIZIA (Artt. 11–15)

L'IA nel lavoro

La legge, nel suo linguaggio programmatico, presenta l'intelligenza artificiale come una risorsa destinata a migliorare le condizioni di vita e di lavoro, a sollevare l'uomo da mansioni ripetitive e ad ampliare la libertà professionale attraverso l'automazione intelligente. Si parla di un uso "etico" e "umanocentrico" dell'IA, capace di rendere il lavoro più sicuro, più efficiente e più equo. Ma dietro questa visione idealizzata si nasconde una realtà molto diversa: nelle pratiche quotidiane, l'intelligenza artificiale non è uno strumento di emancipazione, bensì un **dispositivo di controllo e valutazione** che penetra nelle dinamiche organizzative sostituendo, in maniera invisibile ma capillare, il potere gerarchico tradizionale.

Nelle aziende, nelle piattaforme digitali e perfino nelle pubbliche amministrazioni, gli algoritmi di analisi comportamentale e produttiva vengono impiegati per misurare tempi, prestazioni, spostamenti, interazioni. Ogni attività lascia una traccia digitale che viene registrata, correlata e interpretata come indicatore di efficienza. L'intelligenza artificiale diventa così il nuovo supervisore, un occhio algoritmico che osserva tutto e genera punteggi, classifiche, probabilità di rendimento.

Ciò che un tempo era il giudizio di un superiore oggi è il risultato di un modello predittivo, e la valutazione del lavoratore assume la forma di una *sentenza statistica* che non ammette dialogo né contestazione.

32

Questo mutamento segna un passaggio epocale nella struttura del potere sul lavoro. La sorveglianza non è più verticale, esercitata da un individuo su un altro, ma diffusa e impalpabile, incorporata nelle piattaforme digitali e nei sistemi informativi. L'autorità si sposta dall'essere umano all'algoritmo, e l'obbedienza non si impone più attraverso l'ordine, ma attraverso la **previsione automatizzata**. L'IA non chiede di eseguire un compito: lo anticipa, lo misura e lo valuta in anticipo, stabilendo la soglia di produttività ottimale prima ancora che il lavoratore agisca.

Il potere algoritmico è più pervasivo del potere gerarchico, perché agisce senza visibilità e senza conflitto. Il lavoratore non vede chi lo giudica né può negoziare i criteri della valutazione. Tutto è mediato da modelli di cui non conosce i parametri e da metriche che traducono la complessità umana in numeri. L'oggettività apparente del dato sostituisce la discrezionalità del superiore, ma questa oggettività è illusoria: dietro la neutralità del calcolo si nascondono scelte politiche, economiche e culturali che definiscono cosa sia "produttivo", "performante" o "adatto". L'algoritmo, lungi dall'essere neutro, **incarna un'ideologia della misura** che trasforma la persona in una variabile di efficienza, cancellando il margine di imperfezione che definisce l'umanità del lavoro.

La promessa di miglioramento si traduce così in una nuova forma di disciplinamento. L'IA, anziché alleggerire il carico, lo intensifica: ogni

attività diventa oggetto di tracciamento, ogni pausa viene interpretata, ogni deviazione dai modelli standard è segnale di inefficienza. In nome della produttività, si ricostruisce un regime di controllo costante in cui il lavoratore interiorizza il giudizio della macchina e si adatta ai suoi criteri. È il ritorno del panopticon, ma in forma digitale, dove la visibilità è totale e la consapevolezza nulla.

La legge, nell'invocare un uso etico dell'intelligenza artificiale, non considera questa dimensione di potere. L'etica presuppone un soggetto che possa scegliere, ma l'algoritmo non sceglie: calcola. Parlare di uso "etico" dell'IA in ambito lavorativo significa spostare il problema, poiché l'etica, in un contesto di automazione cognitiva, non può essere affidata a un codice di condotta, ma deve essere iscritta nella struttura stessa dei sistemi, nelle logiche con cui raccolgono, pesano e interpretano i dati.

La sostituzione del potere umano con quello algoritmico modifica radicalmente il contratto sociale del lavoro. Il lavoratore non si confronta più con un capo, ma con un sistema che decide in base a correlazioni invisibili; non si misura con un giudizio, ma con un punteggio. In questo nuovo scenario, la libertà professionale diventa un'apparenza statistica e la dignità del lavoro viene compressa entro le metriche della predizione. L'intelligenza artificiale, che la legge vorrebbe come alleata, assume così il ruolo di **nuovo sovrano del ciclo produttivo**, amministrando in silenzio ciò che un tempo era il frutto del confronto tra persone.

La distanza tra la visione normativa e la realtà tecnologica è dunque totale. Mentre il legislatore auspica un impiego etico, la macchina

costruisce un ecosistema di controllo che riproduce le gerarchie senza nominarle, sostituendo l'autorità con la probabilità. Il potere non comanda più, calcola. E in questo calcolo, l'uomo finisce per essere osservato, classificato e previsto, ma non più realmente compreso.

34

Giustizia e automazione

Nel quadro giuridico attuale, il magistrato viene ancora considerato l'unico soggetto titolare del potere decisionale, custode dell'autonomia di giudizio e garante dell'imparzialità del processo. Tuttavia, questa visione formale non riflette più la realtà dei tribunali contemporanei, dove i sistemi predittivi e le piattaforme di analisi automatizzata stanno progressivamente penetrando nella pratica quotidiana. Gli algoritmi di *risk assessment*, di analisi semantica delle sentenze e di predizione della recidiva vengono utilizzati per orientare le decisioni, suggerire misure cautelari, stimare pene proporzionate o valutare l'attendibilità di prove e testimonianze. In questo contesto, il magistrato non è più un decisore isolato, ma un interprete che opera all'interno di un ambiente cognitivo co-determinato dall'intelligenza artificiale.

Il rischio principale è quello dell'**automation bias**, ovvero la tendenza psicologica a sovrastimare l'affidabilità dei sistemi automatici e a conformarsi alle loro indicazioni. Quando un algoritmo fornisce una valutazione di rischio con una percentuale apparentemente oggettiva — “85% di probabilità di recidiva”, “90% di correlazione con precedenti giurisprudenziali”, “70% di coerenza probatoria” — il giudice si trova di fronte a un numero che si presenta come neutro, razionale e scientifico. Questa cifra, pur derivando da processi probabilistici e dataset

storicamente distorti, esercita un'influenza profonda sul giudizio umano, perché trasmette un senso di precisione e di certezza che la mente tende ad accettare senza analisi critica.

35

La fiducia nell'algoritmo nasce dal suo linguaggio matematico: ciò che appare come calcolo viene percepito come verità. Ma la statistica non è verità, è **una descrizione del passato** travestita da previsione del futuro. I modelli predittivi operano su dati storici che riflettono i bias sociali, economici e giudiziari di epoche precedenti: se un gruppo sociale è stato più spesso oggetto di condanne, l'algoritmo imparerà questa correlazione e la riprodurrà come previsione di rischio. Il magistrato che si affida al sistema senza un controllo critico rischia quindi di **ratificare la discriminazione algoritmica**, amplificando la distorsione sotto la forma dell'imparzialità.

La legge, pur riconoscendo il ruolo di ausilio tecnologico dei sistemi di IA, non affronta la dimensione cognitiva di questa interazione. Manca una riflessione strutturale sulla **formazione dei magistrati rispetto ai bias algoritmici**, e soprattutto non esistono dispositivi giuridici che impongano meccanismi di *de-biasing* obbligatori. Non è previsto, ad esempio, che il giudice debba dimostrare di aver analizzato criticamente le raccomandazioni prodotte dal sistema, né che esista un protocollo per la validazione epistemica delle decisioni algoritmiche. L'assenza di queste misure trasforma la tecnologia da strumento di supporto in **soggetto epistemico non dichiarato**, capace di condizionare il ragionamento giudiziario senza che il diritto ne riconosca la portata.

L'autonomia formale del magistrato, dunque, convive con una dipendenza cognitiva crescente. Nella pratica, il giudice che consulta un

sistema predittivo non si limita a ricevere un consiglio: viene immerso in una logica di probabilità che modella la sua percezione di equità e di rischio. L'atto di giudicare si sposta così dal dominio dell'argomentazione a quello della correlazione statistica, e la giustizia diventa un esercizio di *matching* tra casi, più che un atto di interpretazione dei principi. L'intelligenza artificiale non decide, ma **pre-orienta la decisione**, riducendo progressivamente lo spazio di autonomia intellettuale.

36

Ciò che manca è una presa di coscienza giuridica di questo fenomeno. La legge si limita a prescrivere trasparenza e supervisione umana, ma non affronta la dimensione psicologica del rapporto tra uomo e macchina. In assenza di programmi sistematici di *de-biasing cognitivo*, i magistrati restano esposti alla fascinazione dell'oggettività numerica, convinti di dominare un sistema che, in realtà, orienta le loro scelte. La trasparenza tecnica non basta se non è accompagnata da **alfabetizzazione cognitiva e consapevolezza epistemica**: comprendere come un algoritmo sbaglia è oggi parte integrante del dovere di giudicare.

Il potere giudiziario, senza strumenti per riconoscere e gestire la pressione cognitiva esercitata dai sistemi automatici, rischia di trasformarsi in una forma di giudizio ibrido in cui la legittimità formale resta umana, ma la razionalità effettiva è algoritmica. La fiducia, un tempo riposta nella competenza del giudice, si sposta silenziosamente verso la macchina, e la giustizia diventa un prodotto statistico che parla il linguaggio della certezza, ma si fonda su correlazioni che nessuno controlla più.

CAPITOLO V – DATI, ALGORITMI E METODI MATEMATICI (Artt. 16–17)

La delega al Governo per la disciplina dei dati e degli algoritmi rappresenta uno degli assi portanti della nuova architettura normativa sull'intelligenza artificiale, ma la sua formulazione resta così ampia e generica da compromettere la chiarezza stessa del quadro giuridico che dovrebbe delineare. Si tratta di una delega “aperta”, priva di specificazioni tecniche e di criteri operativi, che attribuisce al potere esecutivo la facoltà di normare ambiti complessi senza definire i confini concettuali fondamentali. La legge non distingue in modo netto tra **algoritmo** e **modello**, confondendo la procedura di calcolo — cioè la sequenza logica che guida un processo — con il risultato dell'apprendimento automatico, ossia la struttura statistica e probabilistica che emerge dall'addestramento su dati reali.

Questa ambiguità semantica non è solo un dettaglio terminologico: produce conseguenze profonde sul piano della responsabilità e della governance. L'algoritmo, inteso come procedura, è un insieme di regole esplicite e verificabili; il modello, invece, è una **entità emergente**, opaca e autonoma, che evolve in funzione del dataset e delle iterazioni di addestramento. Confondere i due significa attribuire alla stessa categoria giuridica soggetti epistemici radicalmente diversi: il primo deterministico e auditabile, il secondo probabilistico e non riproducibile in modo esatto. Senza questa distinzione, non è possibile stabilire chi sia il titolare del trattamento, chi il produttore, chi il responsabile dell'uso e soprattutto **chi risponde dell'errore o della distorsione algoritmica**.

Alla vaghezza concettuale si aggiunge un vuoto infrastrutturale: l'assenza di un **registro pubblico dei dataset di addestramento**. In un'epoca in cui la qualità e la provenienza dei dati determinano la qualità del modello, la mancanza di tracciabilità — il cosiddetto *data lineage* — impedisce di ricostruire la genealogia delle decisioni algoritmiche. Non sapere quali dati siano stati utilizzati, da quali fonti provengano, con quali criteri siano stati selezionati e in che modo siano stati modificati significa **rinunciare a comprendere le cause delle inferenze** prodotte dai sistemi di IA. Senza data lineage, ogni audit è parziale, ogni valutazione etica o scientifica è arbitraria e ogni contestazione legale si riduce a un atto di fede nei confronti dei produttori.

Un modello addestrato su dati non tracciabili diventa un oggetto giuridicamente indeterminato, un'entità senza passato verificabile e senza responsabilità ricostruibile. La trasparenza promessa dalla legge si dissolve nel momento in cui non esiste un archivio che consenta di risalire dal risultato alla fonte, dalla decisione alla conoscenza. Questo vuoto non è solo tecnico, ma politico: impedisce la costruzione di una **sovranità informativa** nazionale, lasciando che la fiducia pubblica si fondi su dichiarazioni di conformità provenienti dagli stessi attori che producono i modelli.

Il nodo più delicato emerge proprio nel contesto geopolitico e industriale: la legge non affronta la questione dei **diritti sui modelli pre-addestrati**, la maggior parte dei quali proviene da soggetti stranieri che detengono la proprietà intellettuale, le chiavi di accesso e gli algoritmi di aggiornamento. Questi modelli — fondamento di gran parte delle applicazioni di IA oggi in uso — sono già strutture chiuse, distribuite in forma di servizio (API o SDK), che non consentono alcun controllo

effettivo sui dati originali né sulle modifiche progressive introdotte durante il fine-tuning. La dipendenza da attori esteri, in assenza di un quadro normativo che ne regoli la proprietà, la responsabilità e i limiti d'uso, espone il Paese a una **perdita di autonomia cognitiva** e a un rischio di vulnerabilità sistemica: i modelli diventano infrastrutture strategiche gestite da entità fuori dalla giurisdizione nazionale.

39

In questa prospettiva, la delega al Governo, priva di criteri sostanziali, appare come una **concessione di potere senza indirizzo tecnico**. La mancanza di una distinzione tra procedure e modelli, l'assenza di un sistema pubblico di tracciabilità dei dataset e la non regolamentazione dei modelli pre-addestrati formano una catena di lacune che mina le fondamenta stesse della sovranità digitale. Senza strumenti per ricostruire l'origine e la vita di un modello, il diritto si trova a legiferare su un oggetto che non può osservare. La legge parla di controllo, ma non fornisce mezzi per esercitarlo. In questa condizione, la fiducia nell'IA diventa un atto di delega cieca, e la sovranità informativa dello Stato si riduce a una forma di dipendenza strutturale da infrastrutture cognitive che appartengono ad altri.

CAPITOLO VI – CYBERSICUREZZA E DIFESA (Art. 18)

L'articolo che promuove la cooperazione tra settore pubblico e privato per il rafforzamento della cybersicurezza nazionale nasce da una visione corretta e necessaria: la sicurezza digitale, oggi, non può più essere garantita da un unico attore. Le infrastrutture critiche, i sistemi sanitari, le reti energetiche e le piattaforme amministrative sono interconnesse e

dipendono in larga misura da tecnologie e servizi forniti da operatori privati. Tuttavia, la formulazione normativa si limita a evocare un principio di collaborazione senza fornire alcuna struttura operativa, nessuna definizione tecnica dei protocolli di interscambio, nessuno standard di interoperabilità né un elenco di soggetti certificatori indipendenti. In assenza di questi elementi, la cooperazione rischia di ridursi a una formula retorica, incapace di produrre un'effettiva architettura di sicurezza condivisa.

40

Il problema è che la cybersicurezza, nella sua accezione moderna, non è più soltanto una questione di protezione dei sistemi informatici, ma di **resilienza cognitiva dell'intero ecosistema informativo**. Le minacce non si limitano a colpire reti e server, ma agiscono sulle menti, sulle percezioni e sulle decisioni delle persone. L'intelligenza artificiale ha introdotto una nuova generazione di vettori d'attacco in grado di sfruttare vulnerabilità non solo tecniche, ma psicologiche e sociali: deepfake altamente realistici, phishing adattivo basato su profili comportamentali, manipolazioni persuasive costruite su modelli linguistici generativi, campagne di sabotaggio cognitivo che colpiscono la fiducia pubblica e distorcono la capacità collettiva di distinguere il vero dal falso.

La legge, nel suo impianto, riconosce la necessità di una collaborazione tra istituzioni e imprese ma ignora che oggi **l'intelligenza artificiale è contemporaneamente difesa e minaccia**, strumento di protezione e di attacco. Senza un quadro tecnico preciso, questa ambivalenza diventa pericolosa. Non basta creare sinergie tra pubblico e privato: serve definire in modo vincolante le regole dell'interoperabilità, le modalità di scambio dei dati di minaccia, i protocolli di risposta coordinata, le

responsabilità nella gestione degli incidenti e, soprattutto, le certificazioni indipendenti dei sistemi di IA utilizzati in contesti di sicurezza nazionale. Senza un'autorità terza e autorevole che possa verificare l'affidabilità dei modelli, ogni cooperazione resta fondata sulla fiducia reciproca, non sulla certezza tecnica.

41

La generica invocazione alla collaborazione omette inoltre di affrontare la questione cruciale della **catena cognitiva della difesa**. Gli attacchi basati sull'intelligenza artificiale non mirano soltanto a interrompere i servizi, ma a modificare la percezione della realtà. Deepfake di leader politici, documenti falsificati con precisione linguistica perfetta, campagne automatizzate di disinformazione possono minare la stabilità democratica più di un'intrusione informatica tradizionale. In questo scenario, la difesa non può limitarsi al perimetro delle reti: deve estendersi alla mente e al comportamento degli individui, creando un sistema di **difesa cognitiva** capace di riconoscere, isolare e neutralizzare le manipolazioni semantiche e percettive prodotte dalle IA ostili.

Una strategia efficace richiederebbe la definizione di standard comuni di *AI security*, la creazione di laboratori nazionali dedicati al testing dei modelli generativi e la costruzione di un'infrastruttura di sorveglianza cognitiva in grado di monitorare le interferenze algoritmiche nei flussi informativi. Ma la norma tace su tutto questo, come se la cybersicurezza fosse ancora un problema di firewall e crittografia, quando invece è ormai un **problema di sovranità cognitiva**. La guerra digitale del XXI secolo si combatte nel linguaggio, nelle immagini e nella fiducia: chi controlla l'intelligenza artificiale controlla la narrazione, e chi controlla la narrazione controlla la percezione della verità.

In assenza di una definizione chiara di ruoli e responsabilità, la cooperazione pubblico-privato rischia di generare una frammentazione del potere difensivo. Ogni attore costruisce la propria infrastruttura, applica protocolli proprietari e mantiene una logica competitiva anziché cooperativa. La mancanza di interoperabilità normativa e tecnica rende impossibile una risposta unitaria agli attacchi basati su IA, che invece operano per natura in modo distribuito e adattivo. Ciò significa che l'aggressore digitale agisce con un'intelligenza collettiva che la legge non è ancora in grado di eguagliare.

Una politica seria di cybersicurezza dovrebbe dunque includere la **formazione cognitiva degli operatori**, la certificazione etica e tecnica dei modelli, la creazione di un'autorità indipendente con competenze multidisciplinari (informatiche, psicologiche, sociologiche e linguistiche), e soprattutto un sistema di trasparenza obbligatoria sui dataset e sugli algoritmi utilizzati nei contesti sensibili. La sicurezza del futuro non sarà una barriera informatica, ma una **rete di consapevolezza** capace di integrare mente umana e macchina difensiva.

La legge, nel suo attuale stato, non riconosce ancora questa transizione. Tratta la cybersicurezza come un problema tecnico, quando è ormai un campo cognitivo e strategico. Senza definizioni chiare, senza protocolli comuni e senza una visione unitaria della difesa cognitiva, la cooperazione tra pubblico e privato rischia di diventare un esercizio di retorica istituzionale, privo della potenza operativa necessaria per affrontare un mondo in cui l'intelligenza artificiale è già, contemporaneamente, il più sofisticato scudo e la più insidiosa delle armi.

CAPITOLO VII – STRATEGIA NAZIONALE E AUTORITÀ (Artt. 19–20)

Governance e istituzioni

La designazione di AgID (Agenzia per l'Italia Digitale) e ACN (Agenzia per la Cybersicurezza Nazionale) come Autorità nazionali per l'intelligenza artificiale appare, a prima vista, una scelta coerente con l'intento di garantire continuità istituzionale e sfruttare strutture già esistenti. Tuttavia, dietro questa decisione si nasconde una fragilità sistemica: nessuna delle due agenzie dispone delle infrastrutture, delle competenze interdisciplinari e delle risorse computazionali necessarie per svolgere una **valutazione scientifica effettiva dei modelli di IA**. La vigilanza, così come delineata dalla norma, rischia di restare puramente teorica, priva degli strumenti operativi indispensabili per analizzare, stressare e certificare i sistemi che sempre più influenzano la vita economica, giuridica e sanitaria del Paese.

AgID, per sua natura, è un ente di indirizzo strategico e regolatorio, non un organismo tecnico-scientifico. La sua missione riguarda la digitalizzazione della pubblica amministrazione, non la valutazione sperimentale di reti neurali o la verifica della robustezza dei modelli di apprendimento. L'ACN, pur avendo competenze più tecniche e un ruolo cruciale nella protezione delle infrastrutture critiche, si concentra sulla **sicurezza informatica**, non sulla sicurezza cognitiva o epistemica dei modelli di IA. La supervisione di algoritmi complessi richiede un livello di competenza che va ben oltre la cybersicurezza: servono fisici, ingegneri, linguisti computazionali, giuristi tecnologici, bioeticisti e

specialisti di *AI alignment*, capaci di misurare non solo le vulnerabilità informatiche, ma anche i bias semantici, le distorsioni statistiche, le derive predittive e gli effetti cognitivi sull'utente umano.

44

L'Italia, al momento, non dispone di un'istituzione pubblica con questo profilo integrato. Mancano **laboratori di test indipendenti**, dotati di accesso a infrastrutture di calcolo ad alte prestazioni (HPC) e di personale formato per condurre *red teaming* algoritmico, audit interpretativi e simulazioni di scenari avversari. In assenza di un *AI Safety Lab* nazionale, la capacità di verifica resta delegata ai produttori stessi, con un evidente conflitto di interessi: chi sviluppa i modelli è anche colui che li certifica. Si tratta di una forma di autoregolamentazione che contraddice l'idea stessa di garanzia pubblica e apre la strada a un regime di sorveglianza nominale, non sostanziale.

L'istituzione di un **Istituto pubblico di validazione algoritmica** rappresenterebbe un passo decisivo per superare questa lacuna. Un ente autonomo, dotato di poteri ispettivi reali e di accesso a risorse HPC, potrebbe diventare il fulcro della sovranità tecnologica italiana in materia di intelligenza artificiale. Il suo compito non dovrebbe limitarsi a verificare la conformità normativa, ma estendersi alla **validazione scientifica e funzionale dei modelli**, attraverso test di robustezza, misurazioni di bias, analisi di trasparenza delle rappresentazioni latenti e prove di resistenza a manipolazioni avversarie. Solo un'infrastruttura di questo tipo può garantire la riproducibilità dei risultati e la tracciabilità dei processi di apprendimento, oggi impossibili da verificare esternamente.

Un laboratorio di sicurezza nazionale per l'IA dovrebbe operare secondo un modello multidisciplinare, simile a quello adottato in altri paesi europei: la Francia ha già istituito il *Laboratoire National pour la Sécurité de l'IA* all'interno dell'INRIA, mentre la Germania dispone di piattaforme coordinate tra DFKI e Fraunhofer. L'Italia, invece, continua a trattare l'intelligenza artificiale come un'estensione del digitale, non come un dominio scientifico autonomo che richiede infrastrutture dedicate e governance specifica. Questa impostazione produce una **asimmetria competitiva** rispetto ai partner europei e rischia di relegare il Paese al ruolo di mero utilizzatore di modelli sviluppati altrove, incapace di verificarne la sicurezza o di influenzarne gli standard.

Un Istituto nazionale di validazione algoritmica, fondato su competenze scientifiche e indipendenza istituzionale, potrebbe restituire alla Repubblica la capacità di **controllare e comprendere** le tecnologie che la governano. Esso dovrebbe disporre di accesso diretto a cluster HPC nazionali, collaborare con le università e con i centri di supercalcolo europei (CINECA, EuroHPC) e fungere da nodo di coordinamento tra enti pubblici, imprese e autorità di garanzia. Attraverso protocolli di audit trasparenti e pubblici, l'Istituto potrebbe pubblicare rapporti periodici sui modelli impiegati in settori sensibili — sanità, giustizia, sicurezza, finanza — restituendo così alla cittadinanza la fiducia nella sorveglianza democratica dell'algoritmo.

In mancanza di una struttura di questo tipo, la vigilanza resta una **finzione burocratica**. AgID e ACN possono definire linee guida, ma non possono validare un modello da miliardi di parametri né verificarne la coerenza interna o il rischio di comportamento emergente. Senza un laboratorio nazionale di sicurezza dell'IA, il Paese si affida alla buona

fede dei produttori e alla capacità dei mercati di autoregolarsi, una scelta che contraddice l'interesse pubblico e la sicurezza nazionale. La vera sovranità digitale non nasce dalle norme, ma dalla **capacità scientifica di ispezionare l'intelligenza**. Solo un Istituto pubblico con poteri ispettivi e mezzi computazionali adeguati potrà colmare questo vuoto, trasformando la vigilanza da principio astratto in garanzia effettiva, e restituendo allo Stato il controllo cognitivo delle proprie infrastrutture digitali.

Sovranità tecnologica

La legge, nel suo impianto dichiarativo, richiama più volte il principio di sovranità tecnologica, presentandolo come fondamento della strategia nazionale sull'intelligenza artificiale. Tuttavia, questa sovranità resta confinata al livello retorico: non vengono previsti investimenti diretti in modelli linguistici nazionali, in dataset pubblici aperti o in infrastrutture di calcolo dedicate. Il risultato è un paradosso: si proclama l'indipendenza tecnologica senza costruirne i presupposti materiali. La sovranità, infatti, non è un atto giuridico, ma una **condizione infrastrutturale e cognitiva**: si fonda sulla capacità di generare, comprendere e controllare i propri modelli di conoscenza, le proprie reti di apprendimento e le proprie basi dati. Senza queste componenti, ogni invocazione di autonomia diventa una forma di dipendenza mascherata.

Oggi l'Italia utilizza modelli linguistici e sistemi di intelligenza artificiale provenienti quasi esclusivamente da fornitori esteri, addestrati su dataset di cui non possiede né la composizione né i diritti. Questa condizione genera una **asimmetria cognitiva**: i modelli che interpretano

la lingua, la scienza e i comportamenti sociali italiani sono plasmatis su epistemologie e priorità culturali non italiane. Le decisioni automatizzate, le previsioni, le classificazioni e perfino i filtri semantici che regolano l'accesso all'informazione rispondono a logiche esterne, a infrastrutture localizzate fuori dal territorio nazionale e a dataset che incorporano valori, metriche e bias di altri contesti geopolitici. In queste condizioni, parlare di sovranità giuridica è un esercizio sterile, perché **non si può essere sovrani di ciò che non si possiede né si comprende.**

L'Italia dovrebbe affrontare la questione della sovranità tecnologica come un progetto scientifico e industriale, non come un principio di policy. Ciò significa creare un ecosistema nazionale di intelligenza artificiale realmente autonomo, fondato su infrastrutture di calcolo ad alte prestazioni, dataset pubblici aperti, e soprattutto su una visione unificata della ricerca. In questo senso, il modello **BRIA** — Bioinformatica, Realtà Immersiva e Intelligenza Artificiale — rappresenta la struttura più avanzata e coerente per costruire un'architettura autonoma di innovazione.

Un modello BRIA nazionale non sarebbe semplicemente un insieme di tecnologie, ma un'infrastruttura cognitiva in grado di integrare le tre grandi dimensioni dell'intelligenza contemporanea: quella biologica, quella digitale e quella esperienziale. La **bioinformatica** fornirebbe il legame con la scienza della vita, con la modellazione dei dati genetici, fisiologici e medici, ponendo le basi per un sistema di IA che rispetti la complessità dell'essere umano e ne protegga la dignità. La **realtà immersiva**, come interfaccia sensoriale e cognitiva, permetterebbe di trasformare i dati in esperienze percepibili, creando ambienti di simulazione, formazione e progettazione collaborativa. L'**intelligenza**

artificiale, infine, agirebbe da motore di analisi, interpretazione e predizione, ma in un quadro regolato da principi etici e scientifici nazionali, non da algoritmi proprietari di soggetti stranieri.

48

Questo modello permetterebbe di costruire un **sistema di innovazione circolare**, in cui la conoscenza prodotta nel Paese alimenta i propri strumenti digitali, e questi, a loro volta, potenziano la ricerca e la formazione interna. Un'infrastruttura BRIA nazionale consentirebbe di creare dataset italiani certificati, costruiti secondo criteri di rappresentatività linguistica, culturale e scientifica, con licenze pubbliche e tracciabilità garantita. In tal modo, la sovranità non si esprimerebbe solo nella capacità di legiferare, ma nella **possibilità di addestrare e validare i propri modelli**, condizione necessaria per ogni indipendenza cognitiva.

L'assenza di un investimento strutturale in questo senso lascia il Paese in una posizione di fragilità. Ogni ministero, ogni istituzione, ogni ente di ricerca dipende da infrastrutture estere per attività essenziali: dall'elaborazione dei dati sanitari alla traduzione automatica, dall'analisi del rischio finanziario alla simulazione industriale. La sovranità tecnologica non può consistere nel semplice "uso conforme" delle tecnologie globali, ma nel loro **governo strutturale**, nella possibilità di aprirle, comprenderle e, se necessario, modificarle.

Occorre dunque passare da una **sovranità normativa a una sovranità epistemica**: la legge deve riconoscere che la vera indipendenza si misura nella capacità di generare conoscenza autonoma, di costruire modelli scientifici proprietari e di custodire i dati come patrimonio

pubblico. Senza questo passaggio, l'Italia continuerà a essere un consumatore digitale in un mondo di produttori cognitivi.

49

Un investimento strategico nel modello BRIA offrirebbe al Paese la possibilità di creare il proprio laboratorio nazionale dell'intelligenza, dove bioinformatica, realtà immersiva e IA lavorano in sinergia per la formazione, la medicina, la difesa, l'energia e la cultura. Sarebbe un **progetto di rinascita tecnologica e civile**, capace di coniugare la visione umanistica italiana con la potenza della scienza contemporanea. La sovranità, in questo senso, non è più soltanto una questione di potere statale, ma di **autonomia cognitiva collettiva**: la capacità di un popolo di comprendere, costruire e difendere i propri strumenti di pensiero.

CAPITOLO VIII – SETTORI APPLICATIVI E PROMOZIONE (Artt. 21–23)

L'articolo 21 rappresenta uno dei pochi passaggi realmente propositivi della legge, in quanto apre alla possibilità di **sperimentazioni ministeriali dirette** nell'ambito dell'intelligenza artificiale. È un segnale positivo perché riconosce, almeno implicitamente, che la pubblica amministrazione non può più limitarsi a regolamentare la tecnologia, ma deve cominciare a utilizzarla, testarla e comprenderla in modo operativo. Tuttavia, il valore di questa misura rischia di restare puramente episodico se non viene inserito in una **visione sistemica di trasformazione amministrativa**. Le sperimentazioni, se isolate e non coordinate, generano una costellazione di progetti paralleli che si

esauriscono nel breve periodo, senza costruire una vera architettura cognitiva pubblica.

50

L'intelligenza artificiale, per la pubblica amministrazione, non deve essere concepita come un insieme di strumenti da collaudare, ma come una **componente strutturale dell'intero apparato statale**, al pari delle infrastrutture digitali, dei registri anagrafici o delle reti energetiche. Ogni ministero, ogni ente territoriale e ogni agenzia dovrebbero poter accedere a un'infrastruttura nazionale comune di IA, condividendo modelli, dati e procedure di validazione. Solo un approccio unitario, capace di trasformare la sperimentazione in governance, può evitare che il Paese si ritrovi con una molteplicità di progetti-pilota privi di interoperabilità, dipendenti da fornitori diversi e incapaci di dialogare tra loro.

Gli **investimenti previsti, pari a un miliardo di euro**, rappresentano una cifra significativa nel contesto europeo, ma rischiano di disperdersi se non accompagnati da una pianificazione strategica coerente. L'allocazione delle risorse dovrebbe seguire una logica di **integrazione verticale**, che parta dall'infrastruttura di calcolo e arrivi fino ai servizi al cittadino, garantendo continuità tra ricerca, formazione, sviluppo e applicazione. Oggi, invece, la mancanza di un piano di lungo periodo rischia di trasformare l'investimento in una sequenza di appalti frammentati, gestiti in modo disomogeneo e incapaci di produrre valore sistemico.

Perché l'articolo 21 produca un effetto reale, è necessario collegarlo a un **piano nazionale di sovranità dei dati**. L'intelligenza artificiale pubblica può esistere solo se lo Stato possiede, custodisce e governa i

propri dati, evitando che la loro gestione venga affidata a soggetti esterni o a infrastrutture collocate fuori dal territorio nazionale. La sovranità dei dati è il fondamento della sovranità decisionale: chi controlla i dati controlla la capacità dello Stato di conoscere, prevedere e decidere. Senza un archivio pubblico unificato, certificato e accessibile per finalità di ricerca e innovazione, ogni sperimentazione ministeriale resta un laboratorio chiuso, scollegato dal sistema cognitivo nazionale.

Accanto alla sovranità dei dati, occorre creare **poli di trasferimento tecnologico** che fungano da cerniera tra la ricerca e la pubblica amministrazione. Questi poli dovrebbero essere sviluppati in stretta connessione con le **ITS Academy e con le Università**, diventando luoghi di formazione avanzata e sperimentazione applicata. Gli ITS possono fornire la base operativa, con tecnici specializzati capaci di gestire e adattare le tecnologie ai contesti amministrativi; le Università, invece, garantiscono la dimensione teorica, etica e giuridica necessaria per interpretare e governare l'evoluzione dei modelli. Un'integrazione stabile tra queste realtà permetterebbe di creare una **filiera nazionale dell'intelligenza pubblica**, capace di trasformare le sperimentazioni in politiche strutturali e durature.

La visione sistemica di cui l'Italia ha bisogno deve superare la logica della "prova" e abbracciare quella della **costruzione organica**: la pubblica amministrazione deve diventare non solo utilizzatrice, ma produttrice di intelligenza artificiale. Ciò significa istituire repository condivisi, standard comuni di progettazione e interoperabilità, protocolli di validazione scientifica e una governance centralizzata che coordini i progetti ministeriali. In assenza di questa regia, ogni sperimentazione

rischia di generare soluzioni proprietarie, frammentate e costose, compromettendo la nascita di un ecosistema coeso e sovrano.

52

L'articolo 21, se letto con spirito lungimirante, potrebbe diventare la base per un **programma nazionale di intelligenza amministrativa**, in cui ogni ministero diventa un laboratorio di innovazione al servizio della collettività, connesso a una rete unitaria di dati, modelli e competenze. Ma perché ciò accada, occorre un atto politico e culturale: riconoscere che l'intelligenza artificiale non è un settore da esplorare, ma una **nuova forma di infrastruttura dello Stato**, tanto necessaria quanto la rete elettrica o quella dei trasporti.

Solo legando le sperimentazioni ministeriali a un piano di sovranità dei dati e a una rete formativa di poli BRIA, ITS e Università, l'Italia potrà trasformare l'investimento di un miliardo in un **progetto di civiltà tecnologica**: un modello in cui l'intelligenza pubblica, la conoscenza scientifica e la formazione professionale convergono per costruire un'amministrazione capace non solo di servire, ma di pensare.

CAPITOLO IX – DELEGHE, FORMAZIONE E FUTURO DEL LAVORO (Art. 24)

Il nucleo più vitale e lungimirante della legge è senza dubbio quello dedicato alla **formazione**, in cui si riconosce per la prima volta il valore dell'"alfabetizzazione algoritmica" come diritto e dovere civico per cittadini e professionisti. È un passaggio epocale, perché implica che la comprensione dell'intelligenza artificiale non sia più un tema riservato agli specialisti, ma una competenza sociale di base, indispensabile per vivere e lavorare in una società sempre più mediata dal calcolo. Tuttavia,

se il principio è innovativo, la sua applicazione concreta mostra limiti evidenti: la formazione delineata è prevalentemente **giuridico-etica**, centrata sulla regolazione dei comportamenti, sulla privacy, sulla deontologia e sulla prevenzione dei rischi. Questa impostazione, pur importante, rischia di produrre una cultura della prudenza senza generare una **competenza reale**. Si educano persone a rispettare norme, ma non a comprendere la natura profonda dei sistemi con cui interagiscono.

53

L'alfabetizzazione algoritmica non può ridursi a un insieme di precetti, deve diventare una **pedagogia cognitiva** capace di connettere il linguaggio della macchina con quello del pensiero umano. Per questo, la formazione in materia di intelligenza artificiale deve essere **tecnico-umanistica**, una disciplina dell'incontro tra logica e filosofia, tra matematica e etica, tra informatica e linguistica. Comprendere l'IA significa capire non solo come funziona, ma *come pensa*: significa riconoscere che ogni rete neurale è una costruzione epistemologica, un modo di rappresentare la realtà e di assegnarle un significato.

L'Italia, con la sua tradizione culturale di equilibrio tra scienza e umanesimo, è nella posizione ideale per costruire questo nuovo paradigma formativo. Le **discipline BRIA — Bioinformatica, Realtà Immersiva e Intelligenza Artificiale — incarnano perfettamente questa sintesi**, perché integrano la conoscenza biologica, la percezione esperienziale e l'intelligenza computazionale in un unico ecosistema educativo. La bioinformatica insegna a leggere il linguaggio della vita attraverso i dati, la realtà immersiva offre un ponte tra l'esperienza sensoriale e il mondo digitale, e l'intelligenza artificiale fornisce gli strumenti per interpretare, predire e creare. Insieme, queste tre

dimensioni costruiscono la **nuova alfabetizzazione del XXI secolo**, dove il sapere non è più compartimentato, ma convergente.

54

La formazione BRIA si distingue proprio per la sua capacità di **trasformare la conoscenza in esperienza**. Un giovane che impara a progettare in realtà immersiva sviluppa la sensibilità percettiva e relazionale; chi studia bioinformatica impara a leggere i codici del corpo e della natura; chi lavora sull'intelligenza artificiale impara a costruire e a interrogare la logica delle decisioni. Questa convergenza produce un pensiero integrato, in cui l'uomo non viene sostituito dalla macchina, ma si ricolloca al suo centro, come progettista del significato, interprete dell'informazione e garante del senso etico della tecnologia.

In assenza di una formazione tecnico-umanistica, l'alfabetizzazione algoritmica rischia di trasformarsi in un semplice adattamento all'automazione, una forma di addestramento passivo alla convivenza con le macchine. Serve invece una **pedagogia della comprensione**, che formi cittadini capaci di leggere gli algoritmi come testi e di decodificare le logiche del potere informazionale che li muovono. Solo chi comprende la grammatica del codice può difendersi dalla sua retorica.

La vera sfida, dunque, non è insegnare a utilizzare l'IA, ma **insegnare a pensare con l'IA**, sviluppando una mente capace di dialogare con l'intelligenza artificiale senza esserne dominata. Questo è il senso profondo di un modello educativo ispirato al paradigma BRIA: un sistema che unisce la precisione della scienza alla profondità della filosofia, la creatività dell'arte alla disciplina dell'ingegneria. In un mondo in cui la macchina pensa, solo una cultura che sappia unire

tecnologia e umanità può garantire la sopravvivenza della libertà intellettuale. L'uomo resterà al centro dell'era digitale non perché la legge lo impone, ma perché la sua formazione gli permetterà di esserlo davvero — consapevole, competente e capace di costruire senso nel cuore stesso del calcolo.

55

CAPITOLO X – CONCLUSIONI GENERALI

La **Legge 132/2025** segna un momento di passaggio cruciale nella storia normativa italiana: è il primo testo che tenta di dare una forma giuridica all'intelligenza artificiale, riconoscendola come fenomeno politico, etico e civile, non più solo tecnologico. In questo senso, è una legge fondativa, un atto di nascita di una nuova stagione della Repubblica digitale. Tuttavia, il suo valore è più simbolico che operativo. Essa non costituisce ancora un corpus di regole tecniche, ma un insieme di enunciazioni di principio, di orientamenti morali e di visioni strategiche. La legge traccia un orizzonte, ma non costruisce ancora la strada per raggiungerlo.

Si tratta di un limite strutturale: il testo parla di tutela, trasparenza e umanocentrismo, ma non definisce **architetture operative** per governare l'intelligenza artificiale nella sua realtà concreta. Manca una visione di sistema che traduca i valori in strumenti, le intenzioni in procedure, i principi in infrastrutture. Il diritto, come spesso accade, arriva con il passo lento delle istituzioni, mentre la tecnologia corre con la velocità esponenziale dell'innovazione. E oggi la distanza tra i due mondi si è fatta vertiginosa.

Le aree in cui la tecnologia ha già superato la legge sono molteplici. La **generatività dei modelli** ha trasformato l'intelligenza artificiale in un'entità creativa, capace di produrre linguaggio, immagini, suoni e decisioni senza supervisione diretta. L'**autonomia decisionale** dei sistemi di machine learning, soprattutto nelle architetture neurali profonde, ha reso impossibile per l'uomo comprendere appieno la catena logica che porta a una determinata inferenza. La **manipolazione dei dati sintetici** ha introdotto una zona grigia tra vero e falso, tra dato e simulazione, dove la stessa nozione di prova giuridica viene messa in crisi. E la **velocità evolutiva delle architetture di apprendimento**, che si aggiornano in modo continuo, rende obsoleti i protocolli di valutazione già nel momento in cui vengono formulati.

Di fronte a questo scenario, la legge, così com'è, non può più limitarsi a sancire principi: deve fondare una **Strategia Nazionale di Sorveglianza Algoritmica**. Tale strategia non dovrebbe avere carattere repressivo, ma conoscitivo, fondato sulla capacità dello Stato di osservare, comprendere e valutare i sistemi in tempo reale. Serve un modello di **audit continuo**, che verifichi non solo la conformità formale, ma anche l'evoluzione dinamica dei modelli impiegati. Occorre un'infrastruttura di **trasparenza open-source**, che consenta alla comunità scientifica e ai cittadini di controllare i comportamenti emergenti delle reti neurali. E deve essere istituito un **registro nazionale dei modelli** utilizzati nella pubblica amministrazione, con indicazione dei dataset di addestramento, dei fornitori, degli aggiornamenti e delle metriche di affidabilità.

Una sorveglianza algoritmica efficace non è una forma di censura, ma una **pratica di democrazia cognitiva**. Essa riconosce che la

conoscenza prodotta dalle macchine non può restare invisibile, che ogni algoritmo che incide sulla vita dei cittadini deve essere tracciabile e interpretabile. Il controllo, in questo contesto, non deve essere esercitato per paura, ma per consapevolezza: la trasparenza è la nuova forma della libertà.

57

Il futuro dell'intelligenza artificiale non sarà determinato dal dominio di una delle due parti — né della legge sulla tecnologia, né della tecnologia sulla legge — ma dalla loro **convivenza**. Il diritto dovrà imparare a dialogare con sistemi che non obbediscono alla logica binaria del sì o del no, ma alla **complessità probabilistica** delle reti neurali, dove la verità non è un valore assoluto, ma una funzione di probabilità. Questo richiederà un profondo rinnovamento epistemologico: i giuristi dovranno formarsi come interpreti della complessità, non come custodi dell'ordine statico.

Solo in questa prospettiva potrà nascere una **giustizia algoritmica** autentica — non una giustizia che teme la macchina, ma che ne comprende la natura e la integra nel proprio orizzonte morale. Una giustizia capace di bilanciare la precisione del calcolo con l'imperfezione della coscienza, la neutralità apparente dell'algoritmo con la responsabilità del giudizio umano. L'uomo resterà il custode del senso non perché la legge lo impone, ma perché nessuna rete, per quanto profonda, potrà mai sostituire la capacità di dare significato. L'intelligenza artificiale costruisce correlazioni, ma solo l'uomo può trasformarle in conoscenza, in discernimento e, infine, in giustizia.