Digeat N.7 - 19 Settembre 2025

Data quality vs. data integrity

Di Paolo Assirelli

Abstract

L'IA sta ridefinendo i paradigmi operativi e decisionali sia nel settore privato che nella pubblica amministrazione, introducendo trasformazioni profonde che interrogano il diritto, l'etica e i modelli organizzativi. Questo articolo tenterà un'analisi dell'implementazione delle tecnologie IA, nello specifico, sulla governance responsabile dei dati attraverso una riflessione sulla sua dipendenza dalla distinzione tra *Data Quality* e *Data Integrity* e sulla necessità di un approccio interdisciplinare che riconosca il ruolo insostituibile del giudizio umano. L'analisi normativa si concentrerà sul quadro regolamentare europeo, includendo il GDPR, il Digital Services Act, l'AI Act e le normative italiane di settore.

Indice

Introduzione

La rapida diffusione dell'IA reca implicazioni profonde, che si estendono ben oltre il mero dominio tecnico, e ha il potenziale di imporre un cambiamento drastico nelle modalità operative delle imprese e delle pubbliche amministrazioni; tuttavia, la fretta di implementare queste tecnologie rischia di mettere in secondo piano l'elemento cardine della gestione approfondita dei dati, così compromettendone il risultato finale.

> La gestione dei dati, infatti, è un prerequisito fondamentale per un'IA responsabile: senza una base dati robusta e ben gestita, i benefici trasformativi promessi dall'IA potrebbero rimanere irrealizzati mentre, al contempo, si amplificano i potenziali rischi operativi, etici e legali.

Il principio ampiamente riconosciuto del "garbage in – garbage out" gioca qui un ruolo fondamentale: dati di bassa qualità conducono a risultati prodotti dagli algoritmi di bassa qualità, ben oltre i semplici errori computazionali, con discriminazioni sistemiche, non conformità legale e una potenziale violazione dei diritti fondamentali.

La discussione sulla Qualità ed Integrità dei Dati, quindi, non è **una semplice buona pratica tecnica, ma una componente critica** per la protezione dei diritti fondamentali nell'era digitale e richiede, in via

preliminare, una corretta distinzione concettuale e applicativa tra *Data Quality* e *Data Integrity*, due costrutti interconnessi ma distinti, entrambi fondamentali per la governance dell'IA e per l'affidabilità dei sistemi che su di essi si basano.

Data Quality e Data Integrity: una distinzione cruciale

Le sfide nell'era dell'IA sono ampliate: i sistemi basati su machine learning richiedono dataset estremamente ampi e di alta qualità, mentre le organizzazioni affrontano carenza di dati adeguati alle restrizioni privacy, costi elevati e deterioramento temporale della rilevanza.

La ***Data Quality*** attiene all'**utilità del dato** per lo scopo previsto. Le dimensioni chiave per la valutazione della qualità dei dati sono molteplici e interconnesse, ciascuna contribuendo alla complessiva affidabilità e usabilità delle informazioni.

Tabella 1 – I pilastri della Data Quality

Criterio	Descrizione
Completezza	Quantità di dati utilizzabili disponibili
Unicità	Assenza di duplicati
Validità	Conformità a formato e regole aziendali
Tempestività	Disponibilità entro i tempi previsti
Accuratezza	Correttezza rispetto a una fonte affidabile o alla realtà
Coerenza	Armonia tra record provenienti da diverse serie di dati o relazioni logiche
Idoneità allo scopo	Soddisfacimento di una specifica esigenza

La ***Data Integrity***, invece, si concentra sulla **sicurezza del dato** e sulla prevenzione di una sua corruzione o alterazione non autorizzata; in altre parole, sulla validità, completezza, coerenza e accuratezza dei dati lungo l'intero ciclo di vita.

L'integrità dei dati è, in effetti, un prerequisito fondamentale per la qualità dei dati. Dati di alta qualità perdono il loro valore se la loro integrità non è garantita, poiché difetti del sistema o errori di trasmissione possono compromettere l'affidabilità dell'intero database.

Tabella 2 – I pilastri della Data Integrity

Criterio	Deacrizione
Autenticità	I dati devono provenire da fonti verificate e attendibili
Tracciabilità	Ogiri modifica ai dati deve essere registrata e attribulbile
Immutabilità	I dati critici non devono essere alterati senza autorizzazione
Sicurezza	I dati devonö essere protetti da accessi non autorizzati s violaźioni
Disponibilità	I dati devono essere accessibili quando necessario, senza interruzioni
Controllo delle versioni	Devono essere gestite correttamete le diverse versioni dei dati nel tempe

> Questa relazione stabilisce una chiara dipendenza gerarchica, dove l'*Integrità dei Dati* funge da strato fondamentale per la *Qualità dei Dati*. Se l'integrità dei dati viene compromessa – sia per accesso non autorizzato, corruzione accidentale o errori sistemici – la loro qualità (accuratezza, coerenza, tempestività) diventa irrilevante o inaffidabile.

Il controllo dell'integrità dei dati deve essere un presupposto imprescindibile per ogni iniziativa basata sui dati, inclusa l'intelligenza artificiale. Affrontare il problema solo a posteriori rende gli interventi di correzione più costosi e complessi, mentre un approccio preventivo garantisce basi solide su cui costruire progetti affidabili. Una governance efficace, dunque, deve assicurare prima la fiducia nei dati e solo dopo puntare a ottimizzarne il valore.

Dati imprecisi o obsoleti non solo perdono utilità, ma si trasformano in un passivo, con conseguenze dirette su ricavi, progetti e capacità decisionale. La frammentazione in silos e la mancanza di standardizzazione aggravano il problema, compromettendo coerenza e uniformità.

Con l'IA, queste criticità si amplificano: sistemi e modelli di machine learning richiedono dataset ampi, consistenti e aggiornati, ma la disponibilità di dati realmente idonei è ostacolata da vincoli di privacy, costi di acquisizione e naturale perdita di rilevanza nel tempo. Se i dati di partenza sono viziati o non rappresentativi, gli algoritmi possono generare distorsioni e propagare bias, minando l'imparzialità delle decisioni. La complessità intrinseca di questi modelli rende inoltre difficile garantire trasparenza e spiegabilità.

In questo contesto, i dati compromessi non solo riflettono i bias preesistenti, ma li amplificano e li diffondono su larga scala, con il rischio di distorsioni sistemiche che minano l'affidabilità complessiva delle soluzioni basate sull'IA.

Caratteristica	Data Quality	Data Integrity
Focus Principale	Utilità dei dati per lo scopo previsto.	Protezione dei dati da alterazioni non autorizzate o accidentali e garanzia della loro coerenza e affidabilità lungo l'intero ciclo di vita.

Obiettivo	Assicurare che i dati siano accurati, completi, validi, unici, tempestivi e coerenti.	Mantenere la validità, completezza, coerenza e accuratezza dei dati attraverso meccanismi di sicurezza e controlli.
Natura	Criteri di valutazione della qualità dei dati.	Misure e tecniche che difendono e salvaguardano i dati.
Dipendenza	Dipende dall'integrità dei dati.	Prerequisito fondamentale per la Data Quality.
Esempi di Problemi	Dati mancanti, duplicati, obsoleti, non formattati, valori incoerenti.	Corruzione dei dati, accessi non autorizzati, perdita di dati, errori di trasferimento, attacchi informatici.
Misure Tipiche	Pulizia dei dati, standardizzazione, deduplicazione, validazione, arricchimento.	Controlli di accesso, crittografia, backup, log di audit, validazione dell'input, controlli transazionali.

Data quality vs. Data Integrity

L'impatto normativo dell'IA

L'approccio europeo si connota per un notevole pragmatismo, concentrandosi su rimedi *ex ante* in chiave di gestione del rischio e approccio *by-design* al prodotto, piuttosto che su rimedi risarcitori *ex post*, considerando che la responsabilità civile tradizionale mal si concilia con la complessità tecnologica e la natura globale degli operatori.

> La trasparenza costituisce un elemento nevralgico della regolamentazione europea, in particolare per i sistemi ad alto rischio. Gli obblighi includono la disclosure dello scopo originale della raccolta dati, dell'uso degli algoritmi e delle informazioni fornite agli utilizzatori.

La normativa richiede inoltre trasparenza per i dataset, pubblici o privati, per garantire la loro "qualità statistica" e mitigare potenziali bias e discriminazioni; l'obbligo di etichettatura per i contenuti sintetici e la tracciabilità del sistema attraverso la "registrazione automatica degli eventi" completano il quadro della trasparenza algoritmica.

L'evoluzione normativa dalla protezione dati tradizionale all'IA si manifesta chiaramente nel passaggio dall'articolo 22 del GDPR, che si concentra sui rischi di trattamenti "unicamente automatizzati" con impatto significativo, all'articolo 86 dell'AI Act, che introduce un diritto assoluto e senza eccezioni alla spiegabilità.

L'AI Act amplia significativamente la prospettiva, richiedendo spiegazioni sul "processo decisionale" e sui "principali elementi della decisione adottata", superando la mera valutazione della qualità dei dati. Questo approccio mira a risolvere il problema della "black-box" algoritmica, rendendo comprensibili i passaggi logici del processo decisionale automatizzato.

Il Quadro Normativo e la Gestione dei Dati nell'IA

Il **GDPR** non menziona esplicitamente il termine “Integrità dei Dati”, tuttavia, le sue disposizioni si ricollegano intrinsecamente sia alla qualità che all’integrità dei dati. I principi cardine del regolamento – **liceità, correttezza e trasparenza del trattamento, limitazione delle finalità, minimizzazione dei dati e accuratezza** – sono intrinsecamente connessi alla data quality.

L’Articolo 7 del GDPR, imponendo **l’adozione di misure di sicurezza appropriate per proteggere i dati personali contro distruzione, perdita, accesso o modificazione non autorizzati, contribuisce direttamente a garantire l’integrità dei dati**, assicurandone affidabilità e consistenza.

> Appare chiaro che l’interazione tra il quadro normativo del GDPR e quello dell’AI Act crei nuove necessità di armonizzazione nella gestione dei dati personali, richiedendo un approccio integrato tale da garantire sia la conformità normativa che la qualità tecnica.

Sotto questo aspetto, l’effettività del consenso si pone come tema di particolare rilevanza, specialmente in relazione ai *cookie* di profilazione e di terze parti. La complessità dei *banner* e dei processi di negazione del consenso, se mal gestiti, potrebbe in pratica svuotare il contenuto del diritto al consenso informato, elemento essenziale nella responsabilità dei sistemi di IA. **La mancanza di disposizioni specifiche sull’adeguatezza delle forme di consenso per l’impiego dei dati nei sistemi di IA rappresenta un fattore di rischio per la violazione dei diritti fondamentali**[1].

Il Considerando 67 dell’AI Act evidenzia esplicitamente l’importanza della **qualità dei dati come “prerequisito essenziale per sistemi di IA affidabili”**. Questa enfasi incide direttamente sulle tecniche di addestramento dei modelli, richiedendo l’adozione di misure appropriate di governance e gestione dei dati.

L’Articolo 10 dell’AI Act dettaglia gli aspetti di governance e gestione dei dati per i sistemi ad alto rischio, includendo le scelte progettuali pertinenti, i processi di raccolta e l’origine dei dati, nonché le operazioni di preparazione quali annotazione, etichettatura, pulizia, aggiornamento, arricchimento e aggregazione[2].

Anche i contratti per lo sviluppo di sistemi di IA necessitano di clausole specifiche che vadano oltre gli accordi *software* tradizionali, includendo dettagli sulla **provenienza, qualità e diritti di utilizzo dei dati di *training***, nonché standard di non discriminazione e livelli di spiegabilità.

Il quadro normativo europeo è arricchito da altri strumenti come il Digital Services Act (DSA) e il Digital Markets Act (DMA), che regolano le piattaforme online, la gestione dei dati e la concorrenza. A questi si aggiungono il Data Governance Act e il Data Act, che completano il sistema di regolamentazione dei dati[3].

> Gli sviluppi normativi europei e nazionali sono un chiaro indice di una progressiva convergenza tra gli aspetti tecnici e quelli giuridici nella *governance* dell’IA, evidenziando la necessità di competenze interdisciplinari per affrontare le sfide della qualità dei dati, della conformità normativa e della responsabilità etica.

L’IA in azienda e nella Pubblica Amministrazione

Le aziende italiane, pur riconoscendo il potenziale dell’IA, si avvicinano a questa tecnologia con cautela, richiedendo concretezza, applicazioni pertinenti al proprio contesto e chiari ritorni sull’investimento. L’implementazione efficace richiede un’attenta gestione dei dati, spesso sottovalutata nelle strategie aziendali.

Un approccio strategico alla Data Quality implica la creazione di team interdisciplinari composti da data architect, esperti di business, data scientist e specialisti della protezione dei dati, a cui demandare la definizione delle aspettative e di obiettivi misurabili tramite indicatori chiave di performance basati sul business.

La *governance* e l'implementazione di processi di *Data Quality* riducono costi e rischi, trasformando i dati in un vero e proprio *asset* competitivo. Ciò richiede lo sviluppo di strutture organizzative che favoriscano una stretta collaborazione tra il *business* e l'IT, promuovendo una cultura aziendale orientata alla *Data Literacy* per coordinare tutti i flussi di gestione delle informazioni, nel rispetto delle normative di settore.

Dal canto suo, **la Pubblica Amministrazione italiana, attraverso le Linee Guida per l'Adozione di IA dell'AGID[4], ha definito modalità specifiche per l'implementazione dei sistemi di IA**, con particolare riferimento alla conformità normativa e all'impatto organizzativo.

Le PA devono prioritariamente rispettare i principi della Carta dei diritti fondamentali dell'UE, la Dichiarazione Europea sui diritti e principi digitali e gli Orientamenti etici per un'IA affidabile dell'AI HLEG[5], promuovendo trasparenza, equità e responsabilità nell'uso dell'IA. Per i sistemi IA classificati ad alto rischio, questi principi sono rafforzati da requisiti specifici definiti dall'AI Act.

> L'AI Act sottolinea la necessità di interventi differenziati di formazione per assicurare un'interazione consapevole con le soluzioni IA. Figure specializzate come l'AI Ethicist, esperti nell'analisi degli impatti sociali dell'IA e nella definizione di linee guida su discriminazione, responsabilità, trasparenza e privacy sono sempre più necessarie.

La gestione della qualità dei dati nella PA è un prerequisito essenziale: le amministrazioni devono condurre un'analisi sistematica dello stato e della qualità dei propri *dataset*, agendo per l'aggiornamento e la pubblicazione come *open data*. L'obiettivo è di rendere i dati interoperabili e integrati per una maggiore efficienza.

Nel settore giudiziario, l'IA ha catalizzato l'attenzione principalmente sul concetto di "giustizia predittiva", termine generico e ambiguo che si riferisce a fenomeni eterogenei: algoritmi basati sull'analisi di precedenti giudiziali tramite machine learning, algoritmi incentrati sull'interpretazione normativa e pratiche virtuose[6] di tribunali che utilizzano banche dati avanzate.

> Esistono limiti epistemici e giuridici significativi nell'applicazione degli algoritmi predittivi nel diritto. Le macchine, pur elaborando massicce quantità di dati a velocità ineguagliabile, mostrano incapacità generale nell'affrontare la complessità del ragionamento giuridico, che trascende il significato letterale delle norme.

Dal punto di vista costituzionale, la sostituzione integrale del giudice con algoritmi decisionali è incompatibile con principi fondamentali quali l'assoggettamento del giudice solo alla legge (Art. 101, 2° comma, Cost.), l'autonomia e l'indipendenza della magistratura, e l'obbligo di motivazione dei provvedimenti giurisdizionali (Art. 111 Cost.).

L'ambito in cui l'IA può parzialmente sostituire il giudice è molto circoscritto, limitandosi alla "giustizia predittiva mite" che riguarda la verifica di requisiti formali, calcoli aritmetici o casi semplici e ripetitivi che escludono l'esercizio della discrezionalità.

Conclusioni

L’adozione dell’intelligenza artificiale e delle tecnologie digitali, tanto nelle imprese quanto nella pubblica amministrazione italiana ed europea, è un processo ormai inevitabile, capace di generare opportunità significative ma anche rischi rilevanti. L’Unione Europea ha assunto un ruolo di guida con un modello regolatorio che pone al centro la dignità umana e i diritti fondamentali, adottando un approccio basato sul rischio e sulla trasparenza, per creare un mercato in cui l’innovazione sia condizionata da valori condivisi.

Il successo di questa trasformazione dipende in misura cruciale dalla qualità e dall’integrità dei dati che alimentano i sistemi di IA. La protezione da alterazioni e la coerenza lungo l’intero ciclo di vita (Data Integrity) sono prerequisiti essenziali per la qualità dei dati, da cui derivano l’accuratezza delle analisi e l’affidabilità delle decisioni automatizzate. Trascurare questi aspetti comporta costi elevati, rischi di bias e perdita di fiducia.

> Il quadro normativo europeo, con GDPR e AI Act, impone requisiti rigorosi per la governance dei dati, la tracciabilità e la qualità dei dataset di addestramento nei sistemi ad alto rischio, riconoscendo un diritto alla spiegabilità più ampio rispetto al GDPR. Tuttavia, restano sfide aperte: frammentazione normativa internazionale, equilibrio tra innovazione e tutela dei diritti, difficoltà di comprendere la logica algoritmica.

Un’adozione responsabile dell’IA richiede team interdisciplinari, formazione continua, codici etici e investimenti in infrastrutture e metodologie di gestione dei dati. Nel settore giudiziario, la cosiddetta “giustizia predittiva” può supportare efficienza e prevedibilità, ma non può sostituire il giudizio umano, né sarebbe compatibile con i principi costituzionali che garantiscono indipendenza e motivazione delle decisioni.

L’IA è uno strumento potente e il suo impatto, positivo o negativo, dipenderà dalla capacità dell’intelligenza umana di governarla. Solo un approccio olistico, che integri tecnologia, processi, diritto e cultura etica dei dati, potrà assicurare che resti al servizio dell’uomo e del bene collettivo. In questo scenario, **il ruolo del giurista** nel “produrre il diritto sempre in meglio” (*cottidie ius in melius produci*), come **“artista della ragione”**[7], rimane un presidio indispensabile nella trasformazione digitale.

NOTE

[1] Si pensi al microtargeting, in particolare quello politico, un meccanismo mediante cui partiti o movimenti politici, attraverso meccanismi di profilazione delle informazioni che gli utenti lasciano sui social network, sono in grado di inviare pubblicità personalizzate agli elettori indecisi e più facili da persuadere. Il GDPR non prevede strumenti adeguati di tutela contro una persuasione raffinata ed indiretta dei potenziali elettori, lasciando ampi margini per la violazione dei diritti connessi al trattamento dei dati personali.

[2] A livello di standardizzazione tecnica, si osserva una chiara evoluzione dagli standard tradizionali ISO/IEC 25012 e ISO/IEC 25024 verso la nuova serie ISO/IEC 5259, specificamente progettata per analytics e machine learning. Il Rapporto Tecnico UNI CEI CEN/CLC/TR 18115-2025 rappresenta il primo documento europeo post-AI Act sulla governance dei dati per l’IA, stabilendo collegamenti tra standard tecnici e requisiti legislativi.

[3] L’Italia, da parte sua, con il disegno di legge sull’IA del 2025, sembra voler introdurre principi specifici per la tutela dei dati personali nell’IA, con attenzione a trasparenza, cybersicurezza e prevenzione delle discriminazioni algoritmiche.

[4] Le ISO/IEC 25012 e 25024 sono riferimenti per la qualità dei dati nelle Linee Guida Open Data di AGID.

[5] High-Level Expert Group on Artificial Intelligence: un gruppo di 52 esperti, scelti dal mondo delle imprese, della ricerca, dell'università e della pubblica amministrazione, che aiuta nell'apertura di un dialogo con tutti gli stakeholder ed è chiamato a proporre analisi e indicazioni per le future politiche europee sull'IA e all'elaborazione delle linee guida sugli aspetti etici, sociali e legali dei sistemi IA.

[6] Esempi di "prassi virtuose" in Italia includono il progetto pilota del Tribunale di Brescia e l'iniziativa della Corte d'Appello di Venezia, che impiegano banche dati evolute per migliorare coerenza e prevedibilità delle decisioni senza sostituire il ruolo del giudice.

[7] G.B. Ferri, Il potere e la parola, in P. LEGENDRE, Il giurista artista della ragione, Torino, 2000.