Digeat N.7 - 19 Settembre 2025

# Le regole invisibili: dalla qualità del dato all’intelligenza normativa

*Di Michele Iaselli*


## Abstract

Questo contributo propone una rilettura critica del concetto di “qualità dei dati”, mostrando come ogni scelta nella costruzione di dataset costituisca un atto normativo con profonde implicazioni sociali. L’articolo introduce il concetto di “intelligenza normativa” per descrivere la capacità emergente dei sistemi artificiali di partecipare attivamente alla produzione e applicazione delle regole sociali. Attraverso l’analisi di casi concreti – dagli algoritmi di hiring ai sistemi di giustizia predittiva – il lavoro dimostra come i dataset funzionino come “costituzioni silenziose” che incorporano specifici sistemi di valori. Ogni decisione di inclusione o categorizzazione dei dati cristallizza principi normativi che influenzeranno migliaia di decisioni algoritmiche future. L’articolo rivela il ruolo del “curatore di dati” come attore normativo e propone lo sviluppo di nuove forme di “gestione responsabile dei dati” basate su trasparenza, inclusività e processi partecipativi. La qualità dei dati diventa così una responsabilità collettiva per garantire che l’intelligenza artificiale rifletta i valori delle società democratiche. Il contributo conclude sostenendo che garantire la qualità dei dati nell’era dell’intelligenza normativa significa garantire la qualità della democrazia stessa, trasformando una sfida tecnica in una questione di cittadinanza digitale.


## Indice


- L’emergere dell’intelligenza normativa
- Il curatore di dati come attore normativo
- La qualità come responsabilità democratica
- La democrazia algoritmica


## L’emergere dell’intelligenza normativa

Nel 2018, Amazon ha dovuto abbandonare il suo sistema di recruiting automatizzato dopo aver scoperto che discriminava sistematicamente le candidate donne[1]. Il problema non risiedeva in un bug del software, ma in qualcosa di più profondo e insidioso: il dataset utilizzato per addestrare l’algoritmo rifletteva dieci anni di decisioni di assunzione in un settore storicamente dominato dagli uomini. **L’intelligenza artificiale aveva semplicemente imparato a perpetuare i bias esistenti, trasformandoli in regole algoritmiche apparentemente oggettive.**

Questo episodio rivela una verità scomoda: quando parliamo di “qualità dei dati”, tendiamo a pensare principalmente a questioni tecniche come completezza, accuratezza e consistenza. Tuttavia, dietro ogni scelta nella costruzione di un dataset si nasconde una dimensione più profonda e spesso invisibile: quella normativa.

> Ogni decisione su cosa includere, come categorizzare, quali metriche privilegiare, non è mai neutra. È un atto di creazione di regole che influenzeranno migliaia, se non milioni, di decisioni future.

**Il caso Amazon illustra perfettamente l'emergere di quella che possiamo definire intelligenza normativa: la capacità dei sistemi artificiali di partecipare attivamente alla produzione, interpretazione e applicazione delle regole sociali**. Non si tratta più di algoritmi che eseguono passivamente istruzioni predefinite, ma di sistemi che, attraverso i pattern appresi dai dati, sviluppano forme autonome di ragionamento pratico con implicazioni dirette per l'azione sociale.

> Questa forma di intelligenza si manifesta quando un algoritmo non si limita a seguire una regola preesistente, ma contribuisce attivamente alla sua definizione o applicazione contestuale.

Accade quando un sistema di valutazione del rischio creditizio non si limita a calcolare probabilità statistiche, ma influenza concretamente la definizione di cosa significhi "affidabilità finanziaria" in quel particolare contesto. Si realizza quando **gli algoritmi di moderazione dei contenuti sui social media non si limitano a rimuovere violazioni evidenti, ma plasmano attivamente le norme comportamentali delle comunità online**, definendo implicitamente cosa sia accettabile e cosa no.

**L'intelligenza normativa** si distingue dall'intelligenza artificiale tradizionale per **tre caratteristiche fondamentali**.

**Primo**, possiede una dimensione autonoma: non si limita a ottimizzare funzioni predefinite, ma sviluppa criteri di valutazione emergenti che possono divergere dalle intenzioni originali dei programmatori.

**Secondo**, presenta una capacità di apprendimento normativo: non impara solo dai dati, ma sviluppa forme di ragionamento pratico che hanno implicazioni dirette per l'azione umana e sociale.

**Infine**, manifesta una dimensione performativa: le sue decisioni non si limitano a descrivere il mondo, ma contribuiscono attivamente a plasmarlo.

Riconoscere l'esistenza dell'intelligenza normativa richiede di andare "oltre il dato" in un senso molto specifico: comprendere che **ogni dataset è una "costituzione silenziosa" che definisce i principi secondo cui l'intelligenza artificiale prenderà le sue decisioni**.

> Come le costituzioni politiche stabiliscono i principi fondamentali di una società, i dataset stabiliscono i principi fondamentali secondo cui un sistema di IA interpreterà e valuterà la realtà.

Un dataset per algoritmi di scoring non contiene semplicemente informazioni finanziarie, ma incorpora una specifica teoria dell'affidabilità creditizia. Se include variabili come il codice postale o la professione, sta codificando l'assunto che il luogo di residenza o il tipo di lavoro siano predittivi della capacità di rimborso. Se esclude certe categorie demografiche per evitare discriminazioni dirette, potrebbe comunque perpetuare bias attraverso proxy variables[2] che correlano indirettamente con caratteristiche protette.

Allo stesso modo, un dataset per sistemi di giustizia predittiva non raccoglie solo dati sui reati: esso, infatti, codifica una particolare visione di cosa costituisca un comportamento "rischioso" o "normale". La scelta di includere arresti e non solo condanne, di considerare reati minori accanto a quelli gravi, di pesare in maniera diversa crimini contro la proprietà e crimini violenti, riflette specifiche concezioni della pericolosità sociale e della filosofia punitiva.

**Un dataset per algoritmi di moderazione dei contenuti non documenta solo violazioni delle regole: materializza specifiche concezioni di cosa sia accettabile nel discorso pubblico**. Le categorizzazioni utilizzate per classificare hate speech, misinformation o spam incorporano visioni particolari della libertà di espressione, del pluralismo democratico e del ruolo delle piattaforme digitali nella sfera pubblica.

## Il curatore di dati come attore normativo

Questa prospettiva ci porta a riconoscere una figura professionale spesso sottovalutata: **il curatore di dati**. Tradizionalmente visto come un ruolo tecnico, il curatore di dati è in realtà **un attore fondamentale nella definizione delle regole che governeranno l'azione degli algoritmi**. Le sue scelte di inclusione ed esclusione, le sue categorizzazioni, le sue metodologie di pulizia e validazione dei dati, rappresentano forme di micropolitica che avranno conseguenze macroscopiche.

> Quando un curatore di dati decide di rimuovere valori "anomali" da un dataset, non sta semplicemente migliorando la qualità statistica dei dati: sta definendo i confini di ciò che il sistema considererà "normale" o "accettabile".

**Un algoritmo addestrato su dati così "puliti" potrebbe poi discriminare sistematicamente comportamenti o caratteristiche che si discostano dalla norma statistica**, anche quando tali deviazioni non sono intrinsecamente problematiche.

Quando sceglie di aggregare certe categorie sociali o demografiche per ragioni di privacy o per semplificare l'analisi, il curatore non sta solo ottimizzando l'efficienza computazionale, ma sta determinando quali distinzioni il sistema sarà in grado di riconoscere e quali invece renderà invisibili. Un sistema sanitario che aggrega troppo genericamente le categorie etniche potrebbe perdere la capacità di rilevare disparità di salute specifiche, perpetuando invisibilmente disuguaglianze esistenti.

Le scelte apparentemente tecniche sulla gestione dei valori mancanti, sulla normalizzazione delle variabili, sulla selezione delle caratteristiche rilevanti, incorporano tutte assunzioni normative sui criteri di rilevanza, sulle modalità appropriate di comparazione, sui compromessi accettabili tra diversi obiettivi.

> **Il curatore di dati opera, quindi, come una sorta di "legislatore ombra" che definisce le regole operative dei sistemi algoritmici**.

## La qualità come responsabilità democratica

Riconoscere la dimensione normativa dei dataset trasforma radicalmente il modo in cui dobbiamo pensare l'imperativo della qualità. La qualità non può più essere definita solo in termini di accuratezza tecnica, ma deve includere dimensioni di giustizia, rappresentatività, trasparenza e accountability. **Un dataset di "alta qualità" non è semplicemente un dataset accurato: è un dataset che riflette consapevolmente i valori che vogliamo vedere incorporati nei sistemi di IA**.

Questo richiede lo sviluppo di nuove metodologie che vadano oltre i tradizionali controlli di qualità tecnica. **Servono audit etici dei dataset che esaminino non solo la correttezza dei dati, ma anche le implicazioni distributive delle scelte di qualità**. È necessario sviluppare analisi degli impatti delle decisioni di classificazione e aggregazione, valutando come diverse scelte metodologiche potrebbero influenzare diversi gruppi sociali.

> La gestione responsabile dei dati deve incorporare meccanismi di partecipazione delle comunità interessate nella definizione dei criteri di qualità. Le persone che saranno soggette alle decisioni algoritmiche dovrebbero avere voce in capitolo nella definizione di cosa costituisca un dataset “rappresentativo” o “equo” per il loro contesto specifico.

Questo non significa sottomettere ogni scelta tecnica al voto popolare, ma **creare canali strutturati di consultazione e feedback che informino le decisioni metodologiche**.

**La trasparenza diventa un principio fondamentale: le scelte di organizzazione dei dati devono essere documentate e giustificate pubblicamente**, non solo negli aspetti tecnici ma anche nelle loro implicazioni etiche e sociali. Ogni dataset dovrebbe essere accompagnato da una “scheda etica” [3] che documenti le decisioni prese, i “compromessi” considerati, i potenziali bias identificati e le misure adottate per mitigarli.

L’inclusività richiede che i processi di costruzione dei dataset includano diverse prospettive disciplinari e sociali. Non basta più assembrare team di soli data scientists: servono competenze interdisciplinari che includano sociologi, filosofi, giuristi, antropologi e, soprattutto, rappresentanti delle comunità che saranno impattate dai sistemi algoritmici.

La revisibilità implica che i dataset devono essere progettati per permettere audit continui e modifiche quando emergono problemi etici o bias sistematici. **I dataset non sono artefatti statici, ma entità viventi che devono poter evolvere al cambiare delle norme sociali e delle comprensioni etiche**.

> L’imperativo della qualità dei dati diventa quindi un imperativo democratico: costruire sistemi di IA che riflettano i valori e gli interessi delle società pluralistiche in cui operano. Non si tratta solo di minimizzare gli errori statistici, ma di massimizzare la giustizia e l’equità delle decisioni algoritmiche.

Questo approccio richiede di ripensare i processi di validazione dei dataset. Accanto ai tradizionali test di accuratezza e precisione, servono metriche di fairness che valutino l’impatto distributivo degli algoritmi su diversi gruppi sociali. Servono procedure di stress testing etico che esaminino come i sistemi si comportano in situazioni limite o con popolazioni marginali.

**La costruzione di dataset di qualità nell’era dell’intelligenza normativa non è più una questione puramente tecnica, ma una forma di diritto in formazione**. Ogni dataset è un frammento di quello che potremmo chiamare “diritto algoritmico”: un sistema di regole incorporate nel codice che governa aspetti sempre più ampi della vita sociale.

## La democrazia algoritmica

Andare “oltre il dato” significa riconoscere questa dimensione normativa e **assumersi la responsabilità di costruire sistemi di IA che non solo funzionino tecnicamente, ma che contribuiscano a una società più giusta ed equa**.

> L’imperativo della qualità diventa, così, un imperativo etico e politico: quello di costruire un futuro in cui l’intelligenza artificiale sia al servizio del bene comune, non solo dell’efficienza algoritmica.

In un’epoca in cui gli algoritmi partecipano sempre più attivamente alla definizione delle regole sociali, garantire la qualità dei dati significa garantire la qualità della democrazia stessa. È una responsabilità che non possiamo permetterci di delegare solo agli esperti, in quanto rappresenta una sfida che

riguarda tutti noi, come cittadini di società che stanno diventando sempre più algoritmiche.

**Le regole invisibili incorporate nei nostri dataset oggi, diventeranno le norme visibili delle nostre società domani**. Spetta a noi decidere che tipo di società vogliamo che queste regole costruiscano.

---

**NOTE**

[1] Jeffrey Dastin, “Amazon scraps secret AI recruiting tool that showed bias against women”, Reuters, 10 ottobre 2018.

[2] Le proxy variables sono variabili che, pur non riferendosi direttamente a caratteristiche protette come razza o genere, correlano statisticamente con esse, permettendo discriminazioni indirette.

[3] Sul modello delle “model cards” proposte da Margaret Mitchell e altri ricercatori di Google per documentare i sistemi di machine learning.