

PROSPETTIVE DALLA PRIMA LINEA

Qualità dei dati, prestazioni dell'AI e fiducia

Data
& Trust
Alliance



Per quanto riguarda l'AI, **la qualità dei dati** si riferisce generalmente all'accuratezza, completezza, coerenza, tempestività, unicità e validità di un determinato set di dati, nonché all'idoneità dei dati allo scopo per il quale vengono utilizzati. Ma il significato pratico di "dati di alta qualità" in un dato contesto dipenderà dalle esigenze dell'organizzazione e dallo specifico caso d'uso coinvolto.

NELL'ECONOMIA DI OGGI, i dati informano di ogni decisione presa dalle aziende, dallo sviluppo del prodotto alla gestione della supply chain, fino alla fatturazione dei clienti. L'incapacità di comprendere e gestire la qualità dei dati può comportare seri rischi a livello legale, finanziari e reputazionali e, in definitiva, limiterà la capacità di un'impresa di innovare e prosperare.

Sostanzialmente, gli algoritmi addestrati su dati errati equivalgono a scarsi risultati aziendali. Dal momento che la maggior parte delle aziende oggi dipende dai dati — sia dati propri sia di quelli a cui accede da terze parti — i leader aziendali devono adottare un approccio misurato e basato sul valore alla qualità dei dati. In questo processo esistono tre fasi fondamentali.



“L'AI si nutre di dati. Se non conosci la qualità dei dati, non potrai conoscere la qualità dei risultati AI finali.”

—JOANN STONIER, ricercatrice di dati e sull'AI presso Mastercard

Queste preoccupazioni sono particolarmente pressanti in un momento in cui i leader aziendali si affrettano ad adottare l'AI sia per l'automazione, sia per l'innovazione che questa tecnologia consente. I modelli AI e gli strumenti aziendali da questi forniti sono efficaci e sicuri quanto i dati sui quali vengono formati e dai quali continuano ad apprendere mentre interagiscono con i clienti e altre fonti di informazione. “L'AI si nutre di dati,” afferma JoAnn Stonier, ricercatrice di dati e sull'AI presso Mastercard. “Se non si conosce la qualità dei dati, non si potrà comprendere la qualità dei risultati AI finali.”

FASE UNO

Stabilisci cosa significa qualità dei dati per la tua azienda

Definire la “qualità” è aspetto altamente contestuale per qualsiasi organizzazione in quanto dipende dagli obiettivi di business e dai casi d'uso di ciascuna azienda. “Nelle varie aziende, ci sono diverse aree di interesse e diversi aspetti della qualità dei dati che sono importanti”, afferma Ioana Mazare, Vice President of Enterprise Data Strategy presso UPS.

Indipendentemente dalle differenze contestuali, tuttavia, ci sono delle considerazioni di base importanti per ogni organizzazione quando si tratta dei dati utilizzati nelle applicazioni AI. Il [World Economic Forum](#) identifica sei elementi fondamentali della qualità dei dati che dovrebbero essere valutati durante la formazione e il controllo dell'AI, sia essa costruita internamente o acquistata da un fornitore:

- **ACCURATEZZA:** puoi confermare (e documentare) che i dati che stai utilizzando rappresentano soggetti reali nel mondo?
- **COMPLETEZZA:** puoi garantire che i dati siano completi e che lo siano anche tutti i relativi valori o campi?

- **COERENZA** : puoi garantire che l'archiviazione dei dati in più luoghi, attraverso reti e applicazioni, sia stabile e coerente nella formattazione e nei valori che rappresenta?
- **TEMPESTIVITÀ**: puoi misurare il ritardo tra il momento in cui i dati vengono generati e il momento in cui vengono utilizzati e garantire che il ritardo non comprometta l'accuratezza dei dati?
- **UNICITÀ**: puoi individuare eventuali duplicazioni o sovrapposizioni tra i set di dati che stai utilizzando per addestrare i tuoi modelli AI?
- **VALIDITÀ**: puoi garantire che i tuoi dati vengano acquisiti nella sintassi o nel formato corretto per lo scopo o per il caso d'uso previsto, inclusi metadati come tipi, intervalli e modelli di dati validi?

dati irrilevanti o duplicati, nonché dati che potrebbero essere stati manomessi, adoperati come "armi" o "avvelenati".

L'idoneità allo scopo riguarda l'ambito, ovvero se i dati rappresentano la popolazione, il mercato o i fattori appropriati per il caso d'uso degli strumenti AI da addestrare. Se i dati che scegli per l'addestramento non sono appropriati allo scopo, l'AI non servirà a nulla. Ad esempio, se un algoritmo di servizi finanziari formato sui dati del mercato europeo viene utilizzato per analizzare i mercati in Africa, il modello non funzionerà in modo efficace.

È importante sottolineare che i set di dati storici possono riflettere pregiudizi storici e un'AI addestrata su tali set di dati perpetuerà tali pregiudizi. "Una parte importante della responsabilità dei dati è garantire di disporre di set di dati completi e abbondanti che



"Nelle varie aziende esistono diverse aree di interesse e diversi aspetti relativi alla qualità dei dati che sono importanti."

—IOANA MAZARE, Vice President of Enterprise Data Strategy presso UPS

Sebbene esistano altri aspetti della qualità dei dati che possono essere importanti per il singolo caso d'uso dell'organizzazione, queste sei considerazioni forniscono un quadro generale delle caratteristiche che contraddistinguono i "dati validi".

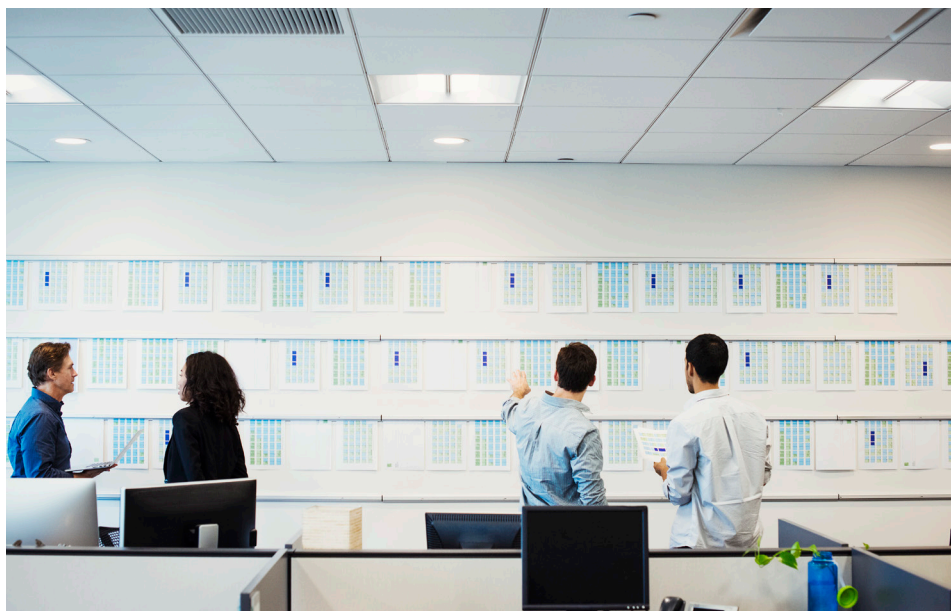
Per garantire la qualità dei dati per AI sono fondamentali due ulteriori dimensioni:

La provenienza dei dati è definita dal [World Wide Web Consortium](#) come le "informazioni su entità, attività e persone coinvolte nella produzione di un dato o di una cosa, che possono essere utilizzati per formare una valutazione sulla sua qualità, affidabilità o attendibilità." La provenienza comprende i dettagli di come la tua azienda ha avuto accesso ai dati e tutti i diritti legali pertinenti, un dato particolarmente importante con i dati di terze parti. La provenienza viene parzialmente acquisita nei [metadati](#) che descrivono in dettaglio l'origine di un pacchetto di dati e le eventuali modifiche apportate ad esso, insieme ai timestamp. Automatizzando la creazione di metadati, le aziende possono aiutare a rilevare ed eliminare

riflettano accuratamente tutte le comunità a cui prestiamo servizio", afferma Manav Misra, Chief Data and Analytics Officer presso Regions Bank. "Quando addestriamo strumenti e modelli AI con una visione veramente olistica, aderiamo all'uso etico e responsabile dei dati e il risultato è una maggiore soddisfazione e inclusione finanziaria per tutti."

Le questioni relative alla provenienza e all'adeguatezza sono una preoccupazione costante per Dena Mendelsohn, Privacy Officer and Compliance Manager presso la startup di assistenza sanitaria digitale Transcarent. "Non dobbiamo entusiasmarci eccessivamente per un set di dati estremamente ampio che apparentemente potrebbe risolvere tutti i nostri problemi", dichiara Mendelsohn. "Dobbiamo scoprire la fonte dei dati, chi li controlla e quali diritti individuali o aziendali si applicano, se ha rilevanza per l'uso previsto e se ci porterà al risultato previsto".

La provenienza dei dati e l'idoneità allo scopo sono particolarmente importanti nel contesto della nuova generazione di AI generativa



pubblicamente disponibile, che è stata addestrata su enormi set di dati da Internet. "Avere la capacità di comprendere la provenienza e la discendenza dei propri dati diventerà sempre più importante per le aziende che desiderano utilizzare l'AI generativa per scopi commerciali", afferma Stonier.

David Cox, Vice President per i modelli AI presso IBM Research, afferma che la provenienza è particolarmente fondamentale. "Ogni leader aziendale dovrebbe chiedersi da dove provengono i propri dati AI e se sono stati puliti e preparati per la formazione", afferma Cox. "Se domani si verifica un problema all'interno di un determinato set di dati, le aziende hanno bisogno dell'impegno affinché i dati vengano rimossi e il modello venga riqualificato."

I principali fornitori di sistemi di data storage, gestione e analytics dei dati tra cui AWS, Azure, Google Cloud e IBM, nonché molte aziende specializzate, offrono strumenti per gestire la provenienza dei dati.

FASE DUE

Creare standard e governance a livello aziendale

La creazione di standard coerenti per la qualità dei dati è fondamentale per un uso etico e responsabile dell'AI. Nel mondo degli affari, tuttavia, non esistono standard universali per la qualità dei dati, dal momento che significa qualcosa di diverso per ogni azienda. Pertanto, è fondamentale che le organizzazioni definiscano i propri standard di qualità dei dati e che lo facciano in modo rigoroso.

Che aspetto ha il "rigore"? Bernardo Tavares, Chief Technology and Data Officer presso Kenvue (ex Johnson & Johnson Consumer Health), ha creato una metrica Data Quality Index (DQI, Indice della qualità dei dati), espressa in percentuale. "Nel caso della qualità dei dati di prodotto, il DQI viene misurato come la percentuale di record disponibili, accurati e collegati in tutti i nostri sistemi", afferma Tavares. "Lo facciamo applicando Business Rules automatizzate. Ci è voluto più di un anno per definire quella metrica." Il suo team ha iniziato con una task force che si è concentrata sui prodotti più redditizi e ha portato i dati a un DQI del 95%.

Oggi, l'azienda è in grado di confrontare i prodotti, da Listerine a Neutrogena, in base alle loro percentuali DQI. "È un numero che monitoriamo a livello di vertici aziendali facile da ricordare e su cui riferire", afferma Tavares. Dopo aver dimostrato l'efficacia del DQI per le parti ad alto valore dell'azienda, il team di Tavares è riuscito a ottenere dei finanziamenti per creare strumenti open source per la creazione e la pulizia dei "record principali" e per iniziare ad espandere la propria iniziativa. "Lentamente ma con decisione, stiamo creando questa nozione di DQI per tutte le nostre entità principali", afferma Tavares.

Gli standard di qualità costituiscono la base della governance dei dati. Il Chief Data Officer (CDO) di un'organizzazione è molto spesso responsabile della creazione di tali standard e dell'implementazione di un framework di governance per applicarli, oltre a collaborare con analisti di dati, data scientist, sviluppatori software, avvocati, team di conformità e chiunque utilizzi strumenti AI in tutta l'azienda.



“Non pretendere l'impossibile. Scegli un'area in cui si trova il punto più critico. Inizia con una domanda chiave a cui non sei in grado di rispondere perché i tuoi dati non sono validi”.

— BERNARDO TAVARES, Chief Technology and Data Officer presso Kenvue

Domande chiave per i vertici aziendali

- Cosa definiamo come **dati mission-critical**?
- Come misuriamo attualmente lo **stato dei nostri dati** per l'AI?
- Valutiamo costantemente i nostri **standard di dati** man mano che le esigenze aziendali cambiano?
- Quale **preparazione e formazione** approntiamo per guidare i dipendenti sull'uso appropriato dei dati?



[Maggiori informazioni su Data & Trust Alliance](#)

[Ottieni insight dagli esperti sull'AI for Business](#)

Secondo un recente [report Gartner](#), i maggiori investimenti in dati e analytics e le richieste poste ai team dati delle organizzazioni, “riflettono una crescente fiducia nelle capacità dei Chief Data and Analytics Officer e il riconoscimento del Data Office come funzione aziendale indispensabile. Tuttavia, ciò comporta più lavoro poiché cresce la pressione affinché il D&A raggiunga risultati aziendali tangibili.

Se la tua organizzazione non dispone di un CDO, non lasciare che questo ti impedisca di agire: come dice Stonier, “è probabile che in azienda ci sia qualcuno che si occupa da tempo della qualità dei dati”. Consigliare al CEO di “esaminare qualsiasi ruolo o funzione esistente con competenza. Potrebbe essere un CIO, qualcuno nel reparto IT o qualcuno nella business intelligence.”

Al di là dei processi specifici di gestione della qualità dei dati di un'azienda, l'impatto sempre più pervasivo di queste nuove funzionalità richiederà la promozione di una cultura incentrata sui dati e sull'AI.

FASE TRE

Creare una cultura aziendale sulla gestione dei dati

“Dovevamo creare una cultura attorno ai dati”, afferma Tavares. “Tutto è iniziato con la dichiarazione che i dati sono un asset reale, e con la comprensione che abbiamo bisogno di persone appassionate in questo ambito e degli strumenti e delle strutture per gestire i dati attraverso i cicli di vita”.

Costruire una cultura dei dati è fondamentale perché il lavoro di governance dei dati non finisce mai. I modelli AI cambieranno continuamente, così come i requisiti normativi, e la governance costante non può essere raggiunta semplicemente creando politiche. I vertici aziendali devono indirizzare gli sforzi di formazione continua sulla qualità dei dati e il CEO deve guidarli. Anche il tono dei vertici, dal CEO in giù, è importante perché gli standard di qualità devono essere integrati in ogni funzione.

Sebbene la costruzione di una cultura aziendale sia un progetto a lungo termine (anzi, senza fine) è probabile che si debba iniziare in piccolo. “Non pretendere l'impossibile”, dice Tavares. Scegli un'area in cui si trova il punto più critico. Inizia con una domanda chiave a cui non sei in grado di rispondere perché i tuoi dati non sono validi.”

Per Tavares, ciò ha portato a concentrarsi sui prodotti dell'azienda, vale a dire là dove gli asset di dati creavano il massimo valore per l'azienda.

Anche questo è stato impegnativo. “Abbiamo riunito nella stanza 70 persone tra esperti e stakeholder e non sono riusciti a mettersi d'accordo su una definizione comune dei dati per un prodotto”, afferma Tavares. “Se ti occupi di ricerca e sviluppo, se ti occupi di vendite, ecc., la vedi in modo diverso.” Il team ha ideato un modello che definisce gli elementi di base e le relazioni che rappresentano davvero un prodotto Kenvue. Solo allora hanno iniziato a lavorare sulla qualità dei dati, “idratando” questo modello con dati puliti e connessi. Aiutato dall'AI, dall'automazione e da un obiettivo di qualità dei dati a livello aziendale, questo approccio iterativo si è rivelato efficace e successivamente è stato applicato a diversi domini “Data360”.

Infine, per costruire una sana cultura attorno ai dati, è fondamentale ricordare che ogni dato rappresenta un essere umano: il tuo cliente, il tuo dipendente, il tuo partner, e la loro presenza e attività nel mondo. Concentrarsi sulle persone nel tuo ecosistema di qualità dei dati (le persone che creano, utilizzano e sono assistite dagli strumenti AI su cui si basano i dati) consentirà alla tua azienda di governare meglio le sue decisioni sull'uso appropriato dei dati nell'AI. “Se progetti pensando all'uomo”, dice Stonier, “i guardrail diventano chiari.” ●