NOVARA
CONSULTING GROUP

This document was translated from English by machine and may contain mistakes. The English original is at https://www.novaracg.com/2025/05/23/four-numbers-and-no-counterfactual-reading-the-artificial-intelligence-case-study-in-human-resources/.

Quattro numeri e nessun controfattuale: lettura del caso di studio sull'intelligenza artificiale nelle Risorse Umane

Heather Grizzle · May 23, 2025

Quattro miglioramenti percentuali e nessun controfattuale. Ciò che ogni metrica standard di un caso di studio sull'IA nelle risorse umane richiede per essere vera, e le domande che richiedono dieci minuti.

Contenuti

1. Il genere
2. Affermazione uno: riduzione del tempo di copertura della posizione
3. Affermazione due: riduzione dell'attività transazionale manuale
4. Affermazione tre: aumento della ritenzione degli alti potenziali
5. Affermazione quattro: mitigazione dei pregiudizi
6. Ciò che manca in tutte e quattro
7. Perché chi lo afferma di solito ci crede
8. Cosa dovrebbe chiedere un acquirente
9. Cosa dovrebbe contenere un caso di studio credibile
10. Conclusione
11. Riferimenti

Il caso di studio sull'intelligenza artificiale nelle risorse umane si è assestato in una forma stabile.

Il genere

Il caso di studio sull'intelligenza artificiale nelle risorse umane si è assestato in una forma stabile. Un'organizzazione descrive un'implementazione, riporta miglioramenti percentuali di tre o quattro punti, aggiunge un'affermazione sulla mitigazione dei pregiudizi e conclude che il talento è diventato una fonte di vantaggio competitivo. I numeri sono specifici. La metodologia è assente. La conclusione invita a una conversazione.

La forma persiste perché funziona come persuasione e perché nessuno è nella posizione di verificarla. I fornitori la pubblicano perché vende. I professionisti la pubblicano perché stabilisce una reputazione. Gli acquirenti la accettano perché l'alternativa sarebbe ammettere di non avere alcuna base per la decisione che stanno per prendere. Ognuno nella catena ha un interesse nel numero e nessuno ha un interesse nel denominatore.

Novara Consulting Group's position is that these documents should be read as advertising claims rather than as evidence, that the specific figures most often cited are the ones most easily produced without any underlying improvement, and that a buyer capable of asking four questions can distinguish a real result from a reported one in about ten minutes. What follows is that method, applied to the four claims that appear most frequently.

Ognuno nella catena ha un interesse nel numero e nessuno ha un interesse nel denominatore.

Affermazione uno: riduzione del tempo di copertura della posizione

Il tempo di copertura della posizione è la metrica più comunemente riportata e quella più facilmente manipolabile senza cambiare nulla nel processo di assunzione.

La misurazione richiede un punto di inizio definito e un punto di fine definito, mantenuti costanti nei periodi di confronto. In pratica il punto di inizio varia a seconda dell'organizzazione e spesso della singola richiesta di assunzione: può essere la data di approvazione della richiesta, la data di pubblicazione dell'annuncio, la data di inizio della ricerca candidati, o la data in cui il responsabile delle assunzioni ha segnalato

per la prima volta l'esigenza. Spostare in avanti questo punto di inizio accorcia l'intervallo misurato senza accorciare nulla di ciò che chiunque sperimenta realmente. Lo stesso vale per l'esclusione delle richieste annullate o ritirate, l'esclusione dei ruoli che hanno richiesto più tempo, o la modifica della composizione dei ruoli nel denominatore.

Il fattore di confondimento è più grande dell'effetto stesso. Il tempo di copertura della posizione risponde fortemente alle condizioni del mercato del lavoro, e i periodi di confronto rilevanti per la maggior parte degli attuali casi di studio coprono un passaggio da una scarsità acuta di manodopera a condizioni considerevolmente più distese. Un'organizzazione che non avesse cambiato nulla mostrerebbe comunque un miglioramento in quell'arco di tempo. Attribuire il movimento a un sistema implementato durante quel periodo richiede un controfattuale, e nessun caso di studio di questo genere ne contiene uno.

Anche la clausola restrittiva merita attenzione. Un miglioramento riportato per i ruoli chiave anziché per tutti i ruoli indica un sottoinsieme, e laddove il sottoinsieme sia stato selezionato dopo che i risultati erano già noti, la cifra descrive la selezione piuttosto che il sistema.

Affermazione due: riduzione dell'attività transazionale manuale

Questa affermazione richiede una misurazione di base dell'attività manuale, e quasi nessuna funzione delle risorse umane ne possiede una.

Where a figure exists it has usually been derived one of three ways: from a vendor's implementation model, from a survey in which staff estimated their own time allocation, or from a count of tickets or transactions that changed definition when the new system was introduced. None of these supports a percentage comparison across time, because the instrument changed at the same moment as the intervention.

There is also a scope question the metric conceals. Transactional work that disappears from the human resources function frequently reappears elsewhere, most often with employees and line managers through self-service. A sixty percent reduction in transactions handled by the function may represent a genuine efficiency gain, a transfer of the same work to people whose time is not being counted, or some combination. Distinguishing these requires measuring total organizational effort, which nobody does,

because the metric exists to justify the function's investment rather than to describe the organization's cost.

Affermazione tre: aumento della ritenzione degli alti potenziali

Questa è la più interessante delle quattro, perché il meccanismo che produce il numero è solitamente l'implementazione stessa.

Alto potenziale non è una caratteristica osservata. È una designazione che l'organizzazione assegna, e un sistema di intelligenza artificiale per i talenti tipicamente cambia il modo in cui viene assegnata. Questo è lo scopo dichiarato. Se il sistema modifica quali dipendenti vengono designati come alto potenziale, allora la coorte misurata dopo l'implementazione è una popolazione diversa dalla coorte misurata prima, e un confronto di ritenzione tra le due non è un confronto della stessa cosa.

La direzione del conseguente bias è prevedibile. Un modello che identifica l'alto potenziale a partire dalla storia delle prestazioni, dai segnali di coinvolgimento e dalla traiettoria di carriera designerà preferenzialmente dipendenti coinvolti, recentemente promossi e in progressione, tutti fattori tra i più forti predittori disponibili della permanenza. Il sistema seleziona persone che sarebbero rimaste comunque e l'organizzazione registra la loro permanenza come un effetto del sistema.

La dimensione della coorte aggrava questo problema. Nella maggior parte delle organizzazioni la popolazione ad alto potenziale è abbastanza piccola che un cambiamento del diciotto percento rappresenta un pugno di individui, ben entro la variazione che qualsiasi piccola coorte produce di anno in anno. Una cifra di questo tipo dovrebbe essere accompagnata dalla dimensione della coorte e dall'indicazione se la definizione sia cambiata. Quasi mai lo è.

Affermazione quattro: mitigazione dei pregiudizi

La quarta affermazione è diversa dalle altre per natura, perché non è semplicemente priva di supporto. È una dichiarazione legale.

Dimostrarla richiede di misurare il pregiudizio prima e dopo, su gruppi definiti, con dimensioni dei campioni adeguate a sostenere il confronto. Richiede anche che i gruppi siano noti, ed è qui che l'affermazione di solito fallisce. La disabilità non è misurabi-

le in questo modo: i datori di lavoro non sanno e non possono legalmente chiedere prima di un'offerta condizionata quali candidati siano disabili, quindi nessuno strumento basato su rapporti d'impatto può registrare una discriminazione per disabilità. Un'organizzazione che afferma di aver mitigato i pregiudizi nella pianificazione della successione ha generalmente misurato razza e sesso, se ha misurato qualcosa, e non ha detto nulla sulla popolazione più colpita dai modelli basati sui dati di prestazione.

La pianificazione della successione comporta una difficoltà ulteriore. Un modello che identifica la prontezza o il potenziale a partire da anzianità di servizio, fase di carriera, tempo nel ruolo, recenza delle credenziali e ritmo delle promozioni precedenti attinge a segnali fortemente correlati con l'età. Riportare la mitigazione dei pregiudizi per un tale sistema, senza averne esaminato la distribuzione per età, afferma l'opposto di ciò che gli input lasciano prevedere.

E l'affermazione stessa crea esposizione. Il Title VII, come emendato nel 1991, rende illegale, in relazione alla selezione o alla segnalazione di candidati per l'impiego o la promozione, modificare i punteggi, usare punteggi minimi diversi o altrimenti alterare i risultati di test correlati all'impiego sulla base di razza, colore, religione, sesso o origine nazionale. Nulla qui suggerisce che questi sistemi facciano ciò. Il punto è che una dichiarazione pubblica che collega uno strumento automatizzato di promozione a esiti demografici è un documento che l'organizzazione ha scritto, non può ritirare, e potrebbe essere chiamata a spiegare.

Ciò che manca in tutte e quattro

Ognuno dei fallimenti precedenti è una versione della stessa assenza. Nessuno di questi casi di studio contiene un controfattuale, ovvero una descrizione di ciò che sarebbe accaduto senza l'intervento.

L'organizzazione ha implementato un sistema durante un periodo in cui il mercato del lavoro è cambiato, l'organico è variato, i manager si sono avvicendati, le retribuzioni sono state adeguate e altre iniziative sono state condotte in parallelo. Ha quindi attribuito il movimento in quattro metriche a uno solo di questi cambiamenti. Questa non è una piccola obiezione metodologica. È l'intera questione, e la sua omissione è ciò che rende il genere infalsificabile.

Le prove pubblicate su affermazioni analoghe dovrebbero far riflettere qualsiasi acquirente. Quando la Harvard Business School e il Burning Glass Institute hanno esami-

nato cosa sia effettivamente accaduto nelle aziende che hanno pubblicamente adottato l'assunzione basata sulle competenze, hanno scoperto che l'eliminazione dei requisiti di laurea ha spostato l'assunzione di candidati senza laurea di circa tre punti e mezzo percentuali, pari a meno di una assunzione su settecento, con quasi la metà delle aziende che non mostrava alcun cambiamento comportamentale nonostante i loro annunci. Gli annunci erano sinceri. La pratica divergeva completamente, e le organizzazioni che li facevano non lo sapevano.

L'inferenza rilevante non è che le persone stiano mentendo. È che le organizzazioni affidabilmente non sanno se le proprie iniziative abbiano funzionato, e che un caso di studio sia scritto dalla parte meno nella posizione di scoprirlo.

Perché chi lo afferma di solito ci crede

Vale la pena affermarlo direttamente, perché la lettura alternativa è poco caritatevole e generalmente sbagliata.

Queste cifre sono raramente fabbricate. Sono prodotte da un ambiente di misurazione in cui non è stata stabilita alcuna base di riferimento, le definizioni sono cambiate, il fornitore ha fornito il quadro analitico, e la persona che compila i risultati è la stessa che ha sostenuto l'acquisto. Ognuna di queste condizioni orienta il risultato nella stessa direzione, e nessuna di esse richiede che qualcuno agisca in malafede. Il numero è reale nel senso che qualcuno lo ha calcolato. Semplicemente non significa ciò che la frase intorno ad esso afferma.

Questo è anche il motivo per cui il contesto normativo ha iniziato a muoversi. La Federal Trade Commission è stata attiva contro le affermazioni non comprovate sulle capacità dell'intelligenza artificiale, inclusa un'ordinanza del gennaio 2025 che ha richiesto a un fornitore di software per l'accessibilità di pagare un milione di dollari per dichiarazioni secondo cui il suo prodotto automatizzato poteva portare qualsiasi sito web in conformità con gli standard di accessibilità. Il principio rilevante non è confinato a quella categoria di prodotto. Le affermazioni sulle prestazioni dell'intelligenza artificiale richiedono una dimostrazione, e un'organizzazione che pubblica percentuali di risultato in un documento di marketing sta facendo esattamente questo tipo di affermazione.

Cosa dovrebbe chiedere un acquirente

Quattro domande, che richiedono circa dieci minuti e alle quali la grande maggioranza dei casi di studio non sopravvive.

Qual era la base di riferimento, quando è stata misurata e da chi. Se la base di riferimento è stata costruita dopo l'implementazione, o fornita dal venditore, il confronto non è tale.

La definizione è cambiata. Per qualsiasi metrica che coinvolga una popolazione designata, come alto potenziale o ruoli chiave, chiedere se i criteri di designazione fossero gli stessi prima e dopo. Se il sistema li ha cambiati, le popolazioni sono diverse e il confronto è nullo.

Cos'altro è cambiato durante il periodo. Mercato del lavoro, organico, retribuzione, gestione, iniziative concorrenti. Un rispondente onesto ne avrà un elenco. Un rispondente senza elenco non ha verificato.

Qual è il denominatore, e quante persone. Le percentuali senza dimensioni della coorte sono inintelligibili, e piccole coorti producono grandi percentuali da piccoli movimenti.

> **UNA QUINTA DOMANDA**
>
> Vale la pena porre una quinta domanda a qualsiasi affermazione che riguardi il bias: quali gruppi sono stati misurati, con quali dimensioni dei campioni, e cosa è stato fatto riguardo ai gruppi che non possono essere misurati. La risposta a quest'ultima parte è quasi sempre che non è stato fatto nulla, il che è di per sé la cosa più utile che l'acquirente imparerà.

Cosa dovrebbe contenere un caso di studio credibile

Il genere non è irrecuperabile, e un'organizzazione disposta a pubblicare quanto segue si distinguerebbe immediatamente in un campo in cui nessuno lo fa.

Indicare il valore di riferimento e la sua data. Indicare le definizioni e segnalare quelle che sono cambiate. Riportare le dimensioni delle coorti insieme alle percentuali. Citare i cambiamenti concomitanti che potrebbero spiegare la variazione, e dire quali non si possono escludere. Descrivere cosa non ha funzionato, perché un'implementazione

in cui tutto è migliorato è un'implementazione in cui non è stato misurato nulla. E laddove venga fatta un'affermazione sul bias, indicare i gruppi, il metodo, le dimensioni dei campioni e ciò che resta non misurato.

Quel documento è meno impressionante della forma attuale e considerevolmente più utile. È anche l'unica versione che sopravvive alla lettura di qualcuno interessato a sapere se è vera.

Conclusione

Il caso di studio è il principale strumento probatorio negli acquisti aziendali di tecnologia per le risorse umane, ed è la forma di prova più debole nell'uso professionale corrente. È riferito a una singola organizzazione, non controllato, autoriportato, definito retrospettivamente, e pubblicato da una parte interessata al risultato.

None of that makes it worthless. It makes it a hypothesis. The organizations that will actually realize returns from these systems are the ones that treat other people's numbers as a reason to investigate rather than a reason to buy, and that hold their own numbers to the standard they would want applied to a vendor's. The discipline is not difficult and it is not expensive. It is a baseline, a stable definition, a cohort size, and an honest account of what else was happening at the time.

Riferimenti

- Federal Trade Commission, *In the Matter of accessiBe, Inc.*, File No. 2223156, denuncia e ordinanza proposta annunciate il 3 gennaio 2025.
- Fuller, Joseph, et al. *Skills-Based Hiring: The Long Road from Pronouncements to Practice*. Harvard Business School Project on Managing the Future of Work and The Burning Glass Institute, febbraio 2024.
- Titolo VII del Civil Rights Act del 1964, come modificato, 42 U.S.C. § 2000e-2, incluso il § 2000e-2(l).
- Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. Part 1607, applicabile alle procedure di selezione che riguardano l'assunzione, la promozione e altre opportunità di impiego.

TIPS

Know something about a sign language AI system that the public should know? Send it to tips@novaracg.com. Every tip is read by me and kept confidential. If you work in the UK, writing to us is not by itself a protected disclosure under the Public Interest Disclosure Act 1998, so before you share employer information, check your options with Protect (protect-advice.org.uk), the free whistleblowing advice line.

Come citare

Grizzle, H. M. (2025, May 23). Quattro numeri e nessun controfattuale: lettura del caso di studio sull'intelligenza artificiale nelle Risorse Umane. Novara Consulting Group. https://www.novaracg.com/it/2025/05/23/four-numbers-and-no-counterfactual-reading-the-artificial-intelligence-case-study-in-human-resources/