NOVARA
CONSULTING GROUP

This document was translated from English by machine and may contain mistakes. The English original is at https://www.novaracg.com/2025/04/04/unpriced-risk-is-not-return-why-most-artificial-intelligence-business-cases-in-human-resources-fail-as-business-cases/.

Il rischio non prezzato non è rendimento: perché la maggior parte dei business case sull'intelligenza artificiale nelle Risorse Umane fallisce come business case

Heather Grizzle · April 4, 2025

The measurable savings in HR AI are small and the large ones cannot be measured. What a defensible AI business case in human resources includes.

Contenuti

1. Le affermazioni e le misurazioni necessarie per sostenerle
2. Ciò che è effettivamente misurabile, e quanto vale
3. La qualità dell'assunzione non è una metrica
4. La previsione dell'abbandono e il divario di accuratezza
5. La conformità automatizzata non estingue un obbligo di conformità
6. I costi che il modello omette
7. Una struttura difendibile
8. Conclusione
9. Riferimenti

Le affermazioni e le misurazioni necessarie per sostenerle

Il modello di business standard per l'intelligenza artificiale nelle risorse umane si è consolidato attorno a cinque affermazioni: assunzioni più rapide e migliori, costi amministrativi inferiori, pianificazione predittiva della forza lavoro, migliore fidelizzazio-

ne e conformità automatizzata. Queste affermazioni vengono ripetute con una coerenza tale nei materiali dei fornitori, nei programmi delle conferenze e nei commenti professionali che ormai funzionano come presupposti di fondo piuttosto che come proposizioni che richiedono prove.

Richiedono prove, e l'ostacolo a produrle non è la tecnologia. Le risorse umane sono entrate in questo periodo con l'infrastruttura di misurazione più debole tra tutte le principali funzioni aziendali. La finanza dispone di bilanci sottoposti a revisione. Le operazioni dispongono di tassi di produttività e di difettosità. Le risorse umane dispongono di un insieme di metriche definite in modo variabile, raramente convalidate, quasi mai dotate di una linea di base prima di un intervento, e spesso prodotte dal fornitore il cui prodotto è oggetto di valutazione. Un'organizzazione che non poteva misurare le prestazioni del proprio processo di assunzione prima di implementare uno strumento automatizzato non può misurarne il miglioramento in seguito, e ciò che riporta come rendimento è generalmente la differenza tra due numeri che non sono mai stati comparabili.

Novara Consulting Group's position is that the central problem with artificial intelligence business cases in human resources is not that the returns are absent. Some are real, and they are described below. It is that returns are asserted at a level of precision the underlying measurement cannot support, while the costs that would offset them are excluded from the model entirely. A calculation constructed that way is not an investment analysis. It is an advocacy document with a currency symbol in it.

Ciò che è effettivamente misurabile, e quanto vale

Le cinque affermazioni standard si suddividono nettamente in due categorie, e questa suddivisione è più istruttiva delle affermazioni stesse.

> È un documento di propaganda con un simbolo di valuta inserito.

La prima categoria contiene effetti genuinamente misurabili e relativamente modesti. Il tempo di ciclo nel reclutamento è misurabile, dispone di una linea di base consolidata nella maggior parte dei sistemi di tracciamento dei candidati, e risponde effettivamente all'automazione della programmazione, della selezione e della corrispondenza. Il costo amministrativo per transazione nella gestione delle buste paga e dei bene-

fit è misurabile, così come i tassi di errore in questi processi. Le organizzazioni che automatizzano queste funzioni generalmente realizzano risparmi, e tali risparmi possono essere documentati.

Due precisazioni sono rilevanti. La prima è che la maggior parte dei risparmi documentati nella gestione delle buste paga e dei benefit deriva dall'automazione dei flussi di lavoro e dall'elaborazione basata su regole, tecnologie che precedono di decenni l'attuale generazione di intelligenza artificiale e che vengono spesso riclassificate come IA nel posizionamento dei fornitori e nei modelli di business costruiti su di esso. Attribuire a un acquisto di IA recente risparmi derivanti da un'automazione dei processi consolidata da tempo gonfia il rendimento misurato di tale acquisto e, cosa più grave, attribuisce erroneamente il meccanismo causale, il che significa che l'organizzazione trae la lezione sbagliata su cosa abbia prodotto il beneficio. La seconda precisazione è che questi risparmi sono, in termini assoluti, poco rilevanti. Non giustificano da soli gli impegni aziendali che vengono attualmente assunti.

La seconda categoria contiene effetti che sarebbero sostanziali se fossero reali e che attualmente non possono essere misurati: migliore qualità delle assunzioni, migliore fidelizzazione e migliore pianificazione della forza lavoro. Sono queste le affermazioni che reggono il modello di business, e lo reggono proprio perché la loro entità non è vincolata da alcuna misurazione.

La qualità dell'assunzione non è una metrica

La qualità dell'assunzione è l'affermazione portante in quasi ogni modello di business relativo alle tecnologie di reclutamento, e non ha una definizione comune convalidata. A seconda dell'organizzazione viene operazionalizzata come valutazione delle prestazioni nel primo anno, fidelizzazione a novanta giorni, fidelizzazione a dodici mesi, soddisfazione del responsabile delle assunzioni, tempo per raggiungere la produttività, oppure un composito di diversi di questi elementi, ciascuno con proprietà di misurazione differenti e nessuno con una validità di costrutto consolidata come misura dell'efficacia delle assunzioni.

La conseguenza è che qualsiasi miglioramento dichiarato nella qualità dell'assunzione è in gran parte un artefatto della scelta definitoria. Un'organizzazione che definisce la qualità dell'assunzione come fidelizzazione a novanta giorni e implementa uno strumento che seleziona candidati simili a dipendenti di lunga data osserverà un miglioramento. Non ha imparato che le assunzioni sono migliorate. Ha imparato che un siste-

ma ottimizzato per un indicatore sostitutivo ha spostato tale indicatore. Quando è il fornitore a fornire sia lo strumento sia il quadro di misurazione, il che è comune, il risultato si avvicina più a una tautologia che a una prova.

Questo rileva ben oltre l'igiene metodologica. Ai sensi delle Uniform Guidelines on Employee Selection Procedures, al 29 C.F.R. Part 1607, un datore di lavoro il cui processo di selezione mostri un impatto sproporzionato deve produrre prove che la procedura sia valida per il ruolo. La validazione basata su criterio richiede una misura difendibile della prestazione lavorativa. Un'organizzazione che non riesce a definire con precisione cosa significhi qualità dell'assunzione per i propri ruoli, e non può dimostrare che la propria misura di criterio sia essa stessa priva di distorsioni, non dispone delle prove previste dal quadro normativo. La debolezza di misurazione che gonfia il modello di business è la stessa debolezza che compromette la difesa legale, un caso insolito in cui un problema metodologico e un problema di responsabilità coincidono.

La previsione dell'abbandono e il divario di accuratezza

La pianificazione predittiva della forza lavoro merita un trattamento separato, perché è qui che la distanza tra le prestazioni pubblicate e quelle effettivamente implementate è maggiore, e l'esposizione legale è meno compresa.

The academic literature on employee attrition prediction routinely reports accuracy figures in the high nineties. Those figures are generated overwhelmingly on a small number of benchmark datasets, principally the IBM HR Analytics set of roughly 1,470 records and a widely circulated Kaggle human resources set of roughly 15,000 records, and typically after synthetic class balancing through oversampling techniques such as SMOTE (Synthetic Minority Over-sampling Technique). These are legitimate methodological choices for comparing algorithms against one another. They are not estimates of field performance on an employer's own workforce, and they should never appear in a business case as though they were. Practitioner assessment of realistic field accuracy for flight risk models is considerably lower, in the range of seventy to eighty percent, and analysts in the field have expressed sustained skepticism toward vendor claims above ninety percent.

Il divario tra l'ottanta per cento e il quasi cento per cento non è una differenza di arrotondamento in questa applicazione, perché il tasso di base è basso. L'abbandono volontario nella maggior parte delle organizzazioni si attesta ben al di sotto del venti per cento annuo, il che significa che un modello con un'accuratezza complessiva elevata

può comunque produrre falsi positivi che superano nettamente i veri positivi. In un modello di abbandono, un falso positivo è un dipendente erroneamente identificato come probabile a lasciare l'azienda.

La questione della governance emerge immediatamente, e l'inquadramento in termini di ritorno sull'investimento la oscura. Cosa fa l'organizzazione con quel nominativo? Se la risposta è nulla, il modello non ha prodotto alcun rendimento. Se la risposta è qualsiasi cosa, che si tratti di offerte di fidelizzazione differenziate, di investimenti di sviluppo modificati, di esclusione da incarichi a lungo termine, o di una diversa considerazione nella successione, allora l'organizzazione ha preso una decisione occupazionale differenziata riguardo a un individuo sulla base di una previsione. Si tratta di un'azione occupazionale, e la sua legittimità dipende da cosa ha determinato il punteggio. Le linee guida per la costruzione di questi modelli raccomandano abitualmente di includere dati demografici dei dipendenti, con età e anzianità citati tra gli input che forniscono contesto al modello. Un datore di lavoro che alimenta un modello con l'età, utilizza l'output per allocare investimenti di sviluppo, e non riesce a ricostruire perché un determinato dipendente ha ricevuto un certo punteggio, ha creato un'esposizione ai sensi dell'Age Discrimination in Employment Act e l'ha registrata nel bilancio come un beneficio.

> **UN RISCHIO OPERATIVO CHE SI AGGRAVA**
>
> Alcuni commentatori hanno anche osservato un rischio operativo che si aggrava e che non compare in nessun modello di business: i punteggi di rischio di abbandono comunicati a responsabili non formati per agire su di essi possono essi stessi provocare la partenza che il modello prevedeva.

La conformità automatizzata non estingue un obbligo di conformità

L'affermazione secondo cui l'intelligenza artificiale garantisce il rispetto della normativa sul lavoro e delle politiche interne dovrebbe essere eliminata del tutto dai modelli di business, e non perché gli strumenti siano inutili. I sistemi di monitoraggio che segnalano pause mancate, soglie di straordinario, anomalie di classificazione o errori nella gestione dei congedi possono essere autenticamente utili, e la motivazione operativa a loro favore è ragionevole.

The word that fails is ensures. A monitoring system does not discharge a legal obligation; it produces outputs the employer then owns. If the system flags a potential violation and no one acts, the organization has created a dated internal record establishing notice, which is materially worse than not having monitored. If the system fails to flag a violation that occurred, the employer has acquired no defense, because reliance on a vendor tool is not a recognized excuse for non-compliance under the Fair Labor Standards Act or the statutes the Equal Employment Opportunity Commission (EEOC) enforces. If the system's determinations are wrong in a patterned way, the employer has automated the error and scaled it.

Gli eventi del gennaio 2025 peggiorano la situazione anziché migliorarla. Quando l'EEOC ha rimosso dal proprio sito web le proprie linee guida sull'intelligenza artificiale, comprese l'assistenza tecnica del maggio 2022 sull'Americans with Disabilities Act e l'assistenza tecnica del maggio 2023 sul Title VII, i datori di lavoro hanno perso il materiale interpretativo che indicava quali pratiche l'agenzia considerasse problematiche. Le leggi non sono cambiate. Uno strumento di conformità configurato sulla base di linee guida nel frattempo ritirate opera ora senza uno standard di riferimento dichiarato, e il datore di lavoro, non il fornitore, è colui che si assume questo vuoto.

L'automazione della conformità aggiunge quindi un obbligo di documentazione e di risposta anziché eliminarne uno. Tale obbligo ha un costo, il costo è reale, e appartiene al lato dei costi del modello.

I costi che il modello omette

La maggior parte dei modelli di business sull'IA nelle risorse umane presenta un lato dei costi costituito da canoni di licenza, servizi di implementazione e una certa previsione per la formazione. Questo è il costo di acquisizione del software. Non è il costo di gestione di un sistema automatizzato di decisioni occupazionali, che comprende inoltre, come minimo, quanto segue.

Il lavoro di validazione, ossia l'analisi che stabilisce che ciò che ogni sistema valuta corrisponde a un requisito documentato del ruolo, insieme alla rivalidazione richiesta quando il sistema o il ruolo cambiano. La gestione del percorso di accomodamento, ossia la capacità permanente di fornire una valutazione alternativa su richiesta, con responsabili decisionali designati, tempi di risposta definiti e un processo interattivo funzionante anziché un semplice modulo. La capacità di conservazione e riproduzione dei registri, ossia la capacità di produrre il punteggio, gli input e la configurazione del

sistema in vigore per qualsiasi decisione individuale durante tutto il periodo di conservazione previsto dal 29 C.F.R. Part 1602. Il monitoraggio, ossia test periodici dei risultati anziché l'affidamento su una sintesi fornita dal fornitore. La cooperazione del fornitore nelle procedure di discovery e nelle indagini amministrative, che è costosa, è assente dai termini commerciali standard a meno che non venga negoziata, e viene addebitata al datore di lavoro quando necessaria. E la funzione di governance interna stessa, ossia la persona identificata e responsabile di quanto sopra, che nella maggior parte delle organizzazioni attualmente non esiste e deve essere creata o affidata in appalto.

Accanto a questi costi operativi si colloca una passività potenziale che praticamente nessun modello di business quantifica, e non è più un'ipotesi. Nell'agosto 2023 l'EEOC ha raggiunto un accordo con iTutorGroup per 365.000 dollari per un software configurato per respingere i candidati al di sopra di età specificate, una cifra concreta per una semplice decisione di configurazione. Nel luglio 2024 il tribunale distrettuale del Northern District of California ha consentito che le azioni legali contro un fornitore di software di selezione procedessero sulla base della tesi secondo cui il fornitore ha agito come agente dei datori di lavoro che utilizzano i suoi strumenti, il che mette in discussione la comune presunzione secondo cui la responsabilità ricada interamente sul datore di lavoro o sul fornitore e possa essere allocata nettamente per contratto. Nel marzo 2025 l'American Civil Liberties Union (ACLU), l'ACLU del Colorado, Public Justice ed Eisenberg and Baum hanno presentato un reclamo alla Colorado Civil Rights Division e all'EEOC per conto di un dipendente Deaf e indigeno riguardo a un colloquio video automatizzato e a una richiesta di accomodamento tramite sottotitolazione negata, una questione che entrambe le aziende negano. Tale pretesa è individuale, non richiede statistiche comparative ed è poco costosa da avanzare.

Il trattamento corretto non consiste nell'ipotizzare una cifra di transazione. Consiste nel dichiarare esplicitamente l'esposizione, nel notare che le pretese relative a pratiche sistemiche sono spesso escluse dalla copertura delle polizze standard di responsabilità per pratiche di lavoro, e nel richiedere che i controlli di governance che riducono l'esposizione siano finanziati come parte dell'implementazione anziché rinviati.

Un'organizzazione che aggiunge queste voci scoprirà spesso che l'investimento ha ancora senso. È un buon risultato e non è questo il punto. Il punto è che la cifra risultante può essere difesa davanti a un consiglio di amministrazione, a un revisore e, eventualmente, a un tribunale, cosa che la cifra attuale non può fare.

Una struttura difendibile

Quanto segue produce un modello di business che resiste al controllo e che può essere costruito da una funzione di risorse umane e finanza già esistente.

1. Stabilire la linea di base prima dell'implementazione, non dopo: tempo di ciclo, costo per assunzione, costo amministrativo per transazione, tassi di errore e abbandono per segmento, ciascuno misurato per almeno quattro trimestri precedenti.
2. Separare i risparmi derivanti dall'automazione da quelli derivanti dall'IA. Indicare quali benefici dichiarati derivano dall'elaborazione dei flussi di lavoro basata su regole che sarebbe disponibile anche senza l'acquisto dell'IA, e quantificarli separatamente.
3. Definire esplicitamente la qualità dell'assunzione per i propri ruoli prima di misurarne qualsiasi variazione, indicare la misura di criterio e indicarne i limiti. Se non può essere definita, non rivendicarne un miglioramento.
4. Refuse benchmark-derived accuracy figures for any predictive model. Require field performance on the organization's own data, with the base rate stated and false positive volume expressed in headcount rather than percentage.
5. Per qualsiasi modello che produca punteggi a livello individuale, documentare quale azione consegue dal punteggio prima dell'implementazione, e far esaminare tale azione come una decisione occupazionale anziché come un output analitico.
6. Quantificare gli obblighi operativi sopra citati come costo ricorrente, non come implementazione una tantum.
7. Dichiarare la passività potenziale come voce informativa distinta, indicando i controlli che la riducono, indipendentemente dal fatto che le venga attribuito un valore in dollari.
8. Richiedere evidenze di validazione, dati sulle prestazioni sul campo e capacità di riproduzione come clausole contrattuali al momento dell'acquisizione. Conservare ogni eventuale rifiuto, che è a sua volta un'informazione su ciò che l'organizzazione sa.

Conclusione

There is a real business case for artificial intelligence in human resources, and it is more modest and more defensible than the one currently in circulation. It rests on cycle time, transaction cost, and error reduction in administrative processes, all of which can be substantiated. It does not rest on quality of hire, retention, or predictive accuracy claims that the function's measurement practice cannot support and that carry liability the model does not price.

Le organizzazioni che riporteranno un ritorno duraturo da questi sistemi non sono quelle che li hanno adottati per prime o che hanno acquistato di più. Sono quelle che hanno misurato prima di implementarli, che possono dire cosa valuta ciascun sistema e perché ciò sia rilevante per il ruolo, e che hanno trattato l'apparato di governance come parte del costo di proprietà piuttosto che come una spesa da evitare. Un ritorno che dipende da un beneficio non misurato e da un rischio non valutato non è un ritorno. È una previsione, e la differenza diventa evidente nel momento in cui qualcuno chiede di vedere i calcoli.

Riferimenti

- American Civil Liberties Union, ACLU of Colorado, Public Justice, e Eisenberg & Baum LLP. Reclamo presentato alla Colorado Civil Rights Division e alla U.S. Equal Employment Opportunity Commission, 19 marzo 2025.
- EEOC v. iTutorGroup, Inc., No. 1:22-cv-02565 (E.D.N.Y.), decreto consensuale, agosto 2023.
- Mobley v. Workday, Inc., No. 3:23-cv-00770 (N.D. Cal.), ordinanza sulla richiesta di archiviazione, luglio 2024.
- Society for Human Resource Management. "The Dangers of Using Predictive Analytics to Gauge Employee Flight Risk." 2023.
- U.S. Equal Employment Opportunity Commission. *The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees*, EEOC-NVTA-2022-2, 12 maggio 2022, e *Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII*, 18 maggio 2023. Entrambi rimossi dal sito web dell'agenzia il 27 gennaio 2025.

- Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. Parte 1607; requisiti di conservazione dei registri, 29 C.F.R. Parte 1602.

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, April 4). Il rischio non prezzato non è rendimento: perché la maggior parte dei business case sull'intelligenza artificiale nelle Risorse Umane fallisce come business case. Novara Consulting Group. https://www.novaracg.com/it/2025/04/04/unpriced-risk-is-not-return-why-most-artificial-intelligence-business-cases-in-human-resources-fail-as-business-cases/