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/.

Unbepreistes Risiko ist keine Rendite: Warum die meisten Business Cases für künstliche Intelligenz im Personalwesen als Business Cases scheitern

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.

Inhalt

1. Die Behauptungen und die Messung, die zu ihrer Stützung erforderlich wäre
2. Was tatsächlich messbar ist, und was es wert ist
3. Einstellungsqualität ist keine Kennzahl
4. Vorhergesagte Fluktuation und die Genauigkeitslücke
5. Automatisierte Compliance entbindet nicht von der Compliance-Pflicht
6. Die Kosten, die das Modell auslässt
7. Eine verteidigungsfähige Struktur
8. Fazit
9. Quellen

Die Behauptungen und die Messung, die zu ihrer Stützung erforderlich wäre

Der Standard-Business-Case für künstliche Intelligenz im Personalwesen hat sich um fünf Behauptungen herum stabilisiert: schnellere und bessere Einstellungen, geringere Verwaltungskosten, vorausschauende Personalplanung, verbesserte Mitarbeiterbindung und automatisierte Compliance. Diese Behauptungen werden in Anbietermaterialien, Konferenzprogrammen und Fachkommentaren mit solcher Konsistenz wiederholt, dass sie inzwischen eher als Hintergrundannahme denn als beweisbedürftige Aussagen fungieren.

Sie erfordern Beweise, und das Hindernis, diese zu erbringen, ist nicht die Technologie. Das Personalwesen ist mit der schwächsten Messinfrastruktur aller großen Unternehmensfunktionen in diese Phase eingetreten. Das Finanzwesen hat geprüfte Abschlüsse. Der Betrieb hat Durchsatz- und Fehlerquoten. Das Personalwesen verfügt über eine Reihe von Kennzahlen, die unterschiedlich definiert, selten validiert, kaum vor einer Intervention als Basiswert erfasst und häufig von genau dem Anbieter erstellt werden, dessen Produkt bewertet wird. Eine Organisation, die die Leistung ihres Einstellungsprozesses vor der Einführung eines automatisierten Tools nicht messen konnte, kann die Verbesserung danach nicht messen, und was sie als Rendite ausweist, ist in der Regel die Differenz zwischen zwei Zahlen, die niemals vergleichbar waren.

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.

Was tatsächlich messbar ist, und was es wert ist

Die fünf Standardbehauptungen lassen sich klar in zwei Klassen unterteilen, und diese Trennung ist aufschlussreicher als die Behauptungen selbst.

> Sie ist ein Plädoyer mit einem Währungssymbol darin.

Die erste Klasse umfasst Effekte, die tatsächlich messbar und vergleichsweise bescheiden sind. Die Durchlaufzeit bei der Personalbeschaffung ist messbar, hat in den meisten Bewerbermanagementsystemen einen etablierten Basiswert und reagiert tatsächlich auf die Automatisierung von Terminplanung, Vorauswahl und Korrespondenz. Die Verwaltungskosten pro Transaktion in der Gehalts- und Sozialleistungsabrechnung sind messbar, ebenso die Fehlerquoten in diesen Prozessen. Organisationen, die diese Funktionen automatisieren, erzielen in der Regel tatsächlich Einsparungen, und diese Einsparungen lassen sich belegen.

Zwei Einschränkungen sind wichtig. Die erste ist, dass die meisten dokumentierten Einsparungen in der Gehalts- und Sozialleistungsabrechnung aus Workflow-Automatisierung und regelbasierter Verarbeitung stammen – Technologien, die der aktuellen Generation künstlicher Intelligenz um Jahrzehnte vorausgehen und in der Anbieterpositionierung sowie in den darauf aufbauenden Business Cases häufig als KI umetikettiert werden. Die Zuschreibung längst etablierter Einsparungen durch Prozessautomatisierung zu einer kürzlich getätigten KI-Anschaffung bläht die gemessene Rendite dieser Anschaffung auf und führt, folgenreicher noch, zu einer Fehlzuordnung des kausalen Mechanismus, was bedeutet, dass die Organisation die falsche Lehre darüber zieht, was den Nutzen tatsächlich erzeugt hat. Die zweite Einschränkung ist, dass diese Einsparungen in absoluten Zahlen wenig spektakulär sind. Sie rechtfertigen allein nicht die unternehmensweiten Verpflichtungen, die derzeit eingegangen werden.

Die zweite Klasse umfasst Effekte, die erheblich wären, wenn sie real wären, und die gegenwärtig nicht gemessen werden können: verbesserte Einstellungsqualität, verbesserte Mitarbeiterbindung und bessere Personalplanung. Dies sind die Behauptungen, die den Business Case tragen, und sie tragen ihn gerade deshalb, weil ihr Umfang durch keine Messung begrenzt wird.

Einstellungsqualität ist keine Kennzahl

Die Einstellungsqualität ist die tragende Behauptung in nahezu jedem Business Case für Rekrutierungstechnologie, und es gibt keine validierte gemeinsame Definition dafür. Je nach Organisation wird sie als Leistungsbeurteilung im ersten Jahr, als Verbleib nach neunzig Tagen, als Verbleib nach zwölf Monaten, als Zufriedenheit der einstellenden Führungskraft, als Zeit bis zur Produktivität oder als Zusammensetzung mehrerer dieser Größen operationalisiert, jeweils mit unterschiedlichen Messeigen-

schaften und ohne etablierte Konstruktvalidität als Maß für die Wirksamkeit der Einstellung.

Die Konsequenz ist, dass jede behauptete Verbesserung der Einstellungsqualität im Wesentlichen ein Artefakt der Definitionswahl ist. Eine Organisation, die Einstellungsqualität als Verbleib nach neunzig Tagen definiert und ein Tool einsetzt, das Kandidaten herausfiltert, die früheren, langjährigen Mitarbeitern ähneln, wird eine Verbesserung beobachten. Sie hat nicht gelernt, dass sich die Einstellung verbessert hat. Sie hat gelernt, dass ein auf einen Stellvertreterwert optimiertes System diesen Stellvertreterwert verschoben hat. Wo der Anbieter sowohl das Tool als auch den Messrahmen liefert, was häufig vorkommt, ähnelt der Befund eher einer Tautologie als einem Beweis.

Dies ist von weit mehr als nur methodischer Bedeutung. Nach den Uniform Guidelines on Employee Selection Procedures gemäß 29 C.F.R. Part 1607 wird von einem Arbeitgeber, dessen Auswahlverfahren eine nachteilige Wirkung zeigt, erwartet, Nachweise zu erbringen, dass das Verfahren für die Stelle valide ist. Eine kriteriumsbezogene Validierung erfordert ein verteidigungsfähiges Maß für die Arbeitsleistung. Eine Organisation, die nicht genau angeben kann, was Einstellungsqualität für ihre eigenen Stellen bedeutet, und die nicht zeigen kann, dass ihr Kriteriumsmaß selbst unvoreingenommen ist, verfügt nicht über die Nachweise, die der Rahmen vorsieht. Die Messschwäche, die den Business Case aufbläht, ist dieselbe Schwäche, die die rechtliche Verteidigung untergräbt – ein ungewöhnlicher Fall, in dem ein methodisches Problem und ein Haftungsproblem ein und dasselbe sind.

Vorhergesagte Fluktuation und die Genauigkeitslücke

Die vorausschauende Personalplanung verdient eine gesonderte Betrachtung, denn hier ist der Abstand zwischen veröffentlichter und tatsächlich erzielter Leistung am größten, und das rechtliche Risiko ist am wenigsten verstanden.

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.

Die Lücke zwischen achtzig Prozent und dem hohen Neunzigerbereich ist in dieser Anwendung kein Rundungsunterschied, weil die Basisrate niedrig ist. Die freiwillige Fluktuation liegt in den meisten Organisationen deutlich unter zwanzig Prozent jährlich, was bedeutet, dass ein Modell mit starker Gesamtgenauigkeit dennoch deutlich mehr falsch-positive als echt-positive Ergebnisse produzieren kann. In einem Fluktuationsmodell ist ein falsch-positives Ergebnis ein Mitarbeiter, der fälschlicherweise als wahrscheinlich kündigend identifiziert wird.

Die Governance-Frage folgt unmittelbar, und die Rahmung als Kapitalrendite verschleiert sie. Was tut die Organisation mit diesem Namen? Lautet die Antwort „nichts“, hat das Modell keine Rendite erbracht. Lautet die Antwort irgendetwas – sei es ein differenziertes Bindungsangebot, veränderte Entwicklungsinvestitionen, der Ausschluss von langfristigen Einsätzen oder eine veränderte Nachfolgeplanung –, dann hat die Organisation eine differenzierende arbeitsrechtliche Entscheidung über eine Person auf Grundlage einer Vorhersage getroffen. Das ist eine arbeitsrechtliche Maßnahme, und ihre Rechtmäßigkeit hängt davon ab, was den Score verursacht hat. Anleitungen zum Aufbau solcher Modelle empfehlen routinemäßig die Einbeziehung von Mitarbeiterdemografie, wobei Alter und Betriebszugehörigkeit als Eingabefaktoren genannt werden, die dem Modell Kontext verleihen. Ein Arbeitgeber, der Alter in ein Modell einspeist, das Ergebnis zur Zuteilung von Entwicklungsinvestitionen nutzt und nicht rekonstruieren kann, warum ein bestimmter Mitarbeiter so bewertet wurde, hat ein Risiko nach dem Age Discrimination in Employment Act geschaffen und es als Nutzen in der Bilanz verbucht.

EIN SICH VERSTÄRKENDES OPERATIVES RISIKO

Kommentatoren haben zudem auf ein sich verstärkendes operatives Risiko hingewiesen, das in keinem Business Case eine Zeile findet: Abwanderungsrisiko-Scores, die an Führungskräfte kommuniziert werden, die nicht darin geschult sind, danach zu handeln, können selbst die Kündigung herbeiführen, die das Modell vorhergesagt hat.

Automatisierte Compliance entbindet nicht von der Compliance-Pflicht

Die Behauptung, künstliche Intelligenz gewährleiste die Einhaltung von Arbeitsrecht und interner Richtlinien, sollte vollständig aus Business Cases gestrichen werden – und nicht, weil die Tools nutzlos wären. Überwachungssysteme, die versäumte Pausen, Überstundengrenzen, Klassifizierungsanomalien oder Fehler in der Urlaubsverwaltung kennzeichnen, können tatsächlich wertvoll sein, und der operative Nutzen dafür ist vernünftig.

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.

Die Ereignisse vom Januar 2025 machen dies eher schlimmer als besser. Als die EEOC ihre Leitlinien zu künstlicher Intelligenz von ihrer Website entfernte, einschließlich der technischen Hinweise vom Mai 2022 zum Americans with Disabilities Act und der technischen Hinweise vom Mai 2023 zu Title VII, verloren Arbeitgeber das Auslegungsmaterial, das ihnen mitteilte, welche Praktiken die Behörde als problematisch ansah. Die Gesetze selbst änderten sich nicht. Ein Compliance-Tool, das gegen inzwischen zurückgezogene Leitlinien konfiguriert wurde, arbeitet nun ohne festgelegten Referenzstandard, und diese Lücke trägt der Arbeitgeber, nicht der Anbieter.

Compliance-Automatisierung fügt daher eine Dokumentations- und Reaktionspflicht hinzu, anstatt eine zu beseitigen. Diese Pflicht verursacht Kosten, diese Kosten sind real, und sie gehören auf die Kostenseite des Modells.

Die Kosten, die das Modell auslässt

Die meisten KI-Business-Cases im Personalwesen enthalten eine Kostenseite, die aus Lizenzgebühren, Implementierungsleistungen und einer gewissen Schulungszulage besteht. Das sind die Anschaffungskosten der Software. Es sind nicht die Betriebskosten eines automatisierten arbeitsrechtlichen Entscheidungssystems, die zusätzlich mindestens Folgendes umfassen.

Validierungsarbeit, das heißt die Analyse, die belegt, dass das, was jedes System bewertet, einer dokumentierten Anforderung der Stelle entspricht, zusammen mit der erforderlichen Revalidierung bei Änderungen des Systems oder der Rolle. Betrieb eines Anpassungswegs, das heißt die ständige Fähigkeit, auf Anfrage eine alternative Bewertung bereitzustellen, mit benannten Entscheidungsträgern, festgelegter Bearbeitungszeit und einem funktionierenden interaktiven Prozess statt eines bloßen Formulars. Fähigkeit zur Aufbewahrung und Reproduktion von Aufzeichnungen, das heißt die Fähigkeit, den Score, die Eingabedaten und die zum Zeitpunkt jeder Einzelentscheidung geltende Systemkonfiguration während des von 29 C.F.R. Part 1602 vorgesehenen Aufbewahrungszeitraums bereitzustellen. Überwachung, das heißt regelmäßige Ergebnisprüfung anstelle des Vertrauens auf eine Anbieterzusammenfassung. Die Mitwirkung des Anbieters bei Beweisaufnahme und behördlicher Untersuchung, die kostspielig ist, in Standard-Geschäftsbedingungen fehlt, sofern nicht gesondert vereinbart, und bei Bedarf dem Arbeitgeber in Rechnung gestellt wird. Und die interne Governance-Funktion selbst, das heißt die benannte Person, die für das Obige verantwortlich ist, die in den meisten Organisationen gegenwärtig nicht existiert und geschaffen oder beauftragt werden muss.

Neben diesen Betriebskosten steht eine Eventualverbindlichkeit, die praktisch kein Business Case beziffert, und sie ist nicht mehr hypothetisch. Im August 2023 einigte sich die EEOC mit iTutorGroup auf 365.000 US-Dollar wegen einer Software, die so konfiguriert war, dass sie Bewerber oberhalb bestimmter Altersgrenzen ablehnte – eine konkrete Zahl für eine einfache Konfigurationsentscheidung. Im Juli 2024 ließ der Northern District of California Klagen gegen einen Anbieter von Screening-Software mit der Begründung zu, der Anbieter habe als Vertreter der Arbeitgeber gehan-

delt, die seine Tools nutzten – was die verbreitete Annahme erschüttert, Haftung liege vollständig entweder beim Arbeitgeber oder beim Anbieter und könne vertraglich klar zugeordnet werden. Im März 2025 reichten die American Civil Liberties Union (ACLU), die ACLU of Colorado, Public Justice und Eisenberg and Baum im Namen eines gehörlosen und indigenen Mitarbeiters eine Beschwerde bei der Colorado Civil Rights Division und der EEOC ein, die ein automatisiertes Videointerview und eine abgelehnte Untertitelungs-Anpassung betraf – ein Sachverhalt, den beide Unternehmen bestreiten. Diese Klage ist individuell, erfordert keine vergleichenden Statistiken und ist kostengünstig zu erheben.

Die richtige Vorgehensweise besteht nicht darin, eine Vergleichssumme zu schätzen. Sie besteht darin, das Risiko ausdrücklich zu benennen, darauf hinzuweisen, dass Klagen wegen systemischer Praktiken häufig außerhalb der Deckung üblicher Employment-Practices-Liability-Policen liegen, und zu verlangen, dass die risikomindernden Governance-Kontrollen als Teil der Einführung finanziert werden, anstatt sie aufzuschieben.

Eine Organisation, die diese Posten hinzufügt, wird häufig feststellen, dass die Investition weiterhin sinnvoll ist. Das ist ein gutes Ergebnis, und darum geht es nicht. Es geht darum, dass die resultierende Zahl vor einem Vorstand, einem Prüfer und schließlich einem Gericht verteidigt werden kann, was die derzeitige Zahl nicht kann.

Eine verteidigungsfähige Struktur

Das Folgende erzeugt einen Business Case, der einer Überprüfung standhält und von einer bestehenden Personal- und Finanzfunktion erstellt werden kann.

1. Legen Sie den Basiswert vor der Einführung fest, nicht danach: Durchlaufzeit, Kosten pro Einstellung, Verwaltungskosten pro Transaktion, Fehlerquoten und Fluktuation nach Segment, jeweils über mindestens vier vorangehende Quartale gemessen.
2. Trennen Sie Automatisierungseinsparungen von KI-Einsparungen. Geben Sie an, welche behaupteten Vorteile aus regelbasierter Workflow-Verarbeitung stammen, die auch ohne die KI-Anschaffung verfügbar wäre, und bewerten Sie diese gesondert.
3. Definieren Sie Einstellungsqualität ausdrücklich für Ihre eigenen Stellen, bevor Sie eine Veränderung darin messen, geben Sie das Kriteriumsmaß an

und benennen Sie dessen Grenzen. Wenn es nicht definiert werden kann, behaupten Sie keine Verbesserung darin.

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. Dokumentieren Sie bei jedem Modell, das Einzelscores erzeugt, vor der Einführung, welche Maßnahme aus dem Score folgt, und lassen Sie diese Maßnahme als arbeitsrechtliche Entscheidung überprüfen, nicht als Analyseergebnis.
6. Bewerten Sie die oben genannten Betriebspflichten als wiederkehrende Kosten, nicht als einmalige Implementierungskosten.
7. Nennen Sie die Eventualverbindlichkeit als ausgewiesenen Posten samt der Kontrollen, die sie mindern, unabhängig davon, ob ihr ein Geldbetrag zugewiesen wird.
8. Verlangen Sie Validierungsnachweise, Felddaten zur Leistung und Reproduzierbarkeit als Vertragsbedingungen beim Erwerb. Bewahren Sie jede Weigerung auf, denn diese ist selbst eine Information darüber, was die Organisation weiß.

Fazit

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.

Die Organisationen, die einen dauerhaften Nutzen aus diesen Systemen berichten werden, sind nicht diejenigen, die am frühesten eingeführt oder am meisten gekauft haben. Es sind diejenigen, die vor dem Einsatz gemessen haben, die sagen können, was jedes System bewertet und warum das für die Stelle von Bedeutung ist, und die den Governance-Apparat als Teil der Gesamtbetriebskosten behandelt haben statt als Ausgabe, die es zu vermeiden gilt. Ertrag, der von ungemessenem Nutzen und unbepreistem Risiko abhängt, ist kein Ertrag. Es ist eine Prognose, und der Unterschied wird genau dann offensichtlich, wenn jemand nach den Berechnungsgrundlagen fragt.

Quellen

- American Civil Liberties Union, ACLU of Colorado, Public Justice und Eisenberg & Baum LLP. Beschwerde eingereicht bei der Colorado Civil Rights Division und der U.S. Equal Employment Opportunity Commission, 19. März 2025.
- EEOC v. iTutorGroup, Inc., No. 1:22-cv-02565 (E.D.N.Y.), Vergleichsbeschluss (Consent Decree) August 2023.
- Mobley v. Workday, Inc., No. 3:23-cv-00770 (N.D. Cal.), Beschluss zum Antrag auf Abweisung, Juli 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. Mai 2022, und *Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII*, 18. Mai 2023. Beide am 27. Januar 2025 von der Website der Behörde entfernt.
- Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. Part 1607; Aufbewahrungspflichten, 29 C.F.R. Part 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.

Zitierweise

Grizzle, H. M. (2025, April 4). Unbepreistes Risiko ist keine Rendite: Warum die meisten Business Cases für künstliche Intelligenz im Personalwesen als Business Cases scheitern. Novara Consulting Group. https://www.novaracg.com/de/2025/04/04/unpriced-risk-is-not-return-why-most-artificial-intelligence-business-cases-in-human-resources-fail-as-business-cases/