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/.

Quatre chiffres et aucun contrefactuel : lecture de l'étude de cas sur l'intelligence artificielle en ressources humaines

Heather Grizzle · May 23, 2025

Quatre améliorations en pourcentage et aucun contrefactuel. Ce que chaque indicateur type d'étude de cas sur l'IA en RH exige comme vérité, et les questions qui prennent dix minutes.

Sommaire

1. Le genre
2. Affirmation un : réduction du délai de recrutement
3. Affirmation deux : réduction de l'activité transactionnelle manuelle
4. Affirmation trois : augmentation de la rétention des hauts potentiels
5. Affirmation quatre : atténuation des biais
6. Ce qui manque dans les quatre cas
7. Pourquoi le demandeur y croit généralement
8. Ce qu'un acheteur devrait demander
9. Ce que contiendrait une étude de cas crédible
10. Conclusion
11. Références

L'étude de cas sur l'intelligence artificielle dans les ressources humaines s'est installée dans une forme stable.

Le genre

L'étude de cas sur l'intelligence artificielle dans les ressources humaines s'est installée dans une forme stable. Une organisation décrit un déploiement, rapporte trois ou quatre points de pourcentage d'amélioration, ajoute une affirmation sur l'atténuation des biais, et conclut que le talent est devenu une source d'avantage concurrentiel. Les chiffres sont précis. La méthodologie est absente. La conclusion invite à la conversation.

La forme persiste parce qu'elle fonctionne comme un argument de persuasion et parce que personne n'est en position de la vérifier. Les fournisseurs la publient parce qu'elle se vend. Les praticiens la publient parce qu'elle établit une réputation. Les acheteurs l'acceptent parce que l'alternative serait d'admettre qu'ils n'ont aucune base pour la décision qu'ils s'apprêtent à prendre. Chaque maillon de la chaîne a intérêt au chiffre et aucun n'a intérêt au dénominateur.

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.

Chaque maillon de la chaîne a intérêt au chiffre et aucun n'a intérêt au dénominateur.

Affirmation un : réduction du délai de recrutement

Le délai de recrutement est la mesure la plus couramment rapportée et la plus facile à faire bouger sans rien changer à l'embauche.

La mesure requiert un point de départ défini et un point d'arrivée défini, maintenus constants sur les périodes comparées. Dans la pratique, le point de départ varie selon l'organisation et souvent selon le poste à pourvoir : il peut s'agir de la date d'approbation du poste, de la date de publication de l'offre, de la date de début du sourcing, ou de la date à laquelle le responsable du recrutement a initialement exprimé le besoin. Déplacer ce point de départ plus tard raccourcit l'intervalle mesuré sans raccourcir

quoi que ce soit de vécu par qui que ce soit. Il en va de même pour l'exclusion des postes annulés ou retirés, l'exclusion des postes ayant pris le plus de temps, ou le changement de la composition des postes dans le dénominateur.

Le facteur de confusion est plus important que l'effet. Le délai de recrutement réagit fortement aux conditions du marché du travail, et les périodes de comparaison pertinentes pour la plupart des études de cas actuelles couvrent une transition d'une pénurie aiguë de main-d'œuvre vers des conditions nettement plus souples. Une organisation qui n'aurait rien changé montrerait une amélioration sur cette période. Attribuer ce mouvement à un système déployé pendant cette période nécessite un contrefactuel, et aucune étude de cas de ce genre n'en contient.

Le qualificatif mérite également attention. Une amélioration rapportée pour les postes clés plutôt que pour tous les postes indique un sous-ensemble, et lorsque ce sous-ensemble a été sélectionné après que les résultats étaient connus, le chiffre décrit la sélection plutôt que le système.

Affirmation deux : réduction de l'activité transactionnelle manuelle

Cette affirmation requiert une mesure de référence de l'activité manuelle, et presque aucune fonction ressources humaines n'en dispose.

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.

Affirmation trois : augmentation de la rétention des hauts potentiels

C'est la plus intéressante des quatre, car le mécanisme qui produit le chiffre est généralement le déploiement lui-même.

Le haut potentiel n'est pas une caractéristique observée. C'est une désignation que l'organisation attribue, et un système d'intelligence artificielle dédié aux talents modifie généralement la façon dont elle est attribuée. C'est son objectif déclaré. Si le système modifie quels employés sont désignés comme hauts potentiels, alors la cohorte mesurée après déploiement est une population différente de celle mesurée avant, et une comparaison de rétention entre les deux ne compare pas la même chose.

Le sens du biais qui en résulte est prévisible. Un modèle identifiant le haut potentiel à partir de l'historique de performance, des signaux d'engagement et de la trajectoire de carrière désignera préférentiellement des employés engagés, récemment promus et en progression, autant de facteurs qui comptent parmi les prédicteurs les plus solides de la fidélisation. Le système sélectionne des personnes qui allaient rester, et l'organisation enregistre leur maintien en poste comme un effet du système.

La taille de la cohorte aggrave ce phénomène. Dans la plupart des organisations, la population des hauts potentiels est suffisamment restreinte pour qu'un changement de dix-huit pour cent représente une poignée d'individus, bien à l'intérieur de la variation que toute petite cohorte produit d'une année sur l'autre. Un chiffre de ce type devrait être accompagné de la taille de la cohorte et de l'indication de savoir si la définition a changé. Ce n'est presque jamais le cas.

Affirmation quatre : atténuation des biais

La quatrième affirmation diffère des autres par nature, car elle n'est pas simplement non étayée. C'est une déclaration à portée juridique.

L'étayer requiert de mesurer le biais avant et après, sur des groupes définis, avec des effectifs de cellules suffisants pour soutenir la comparaison. Cela requiert également que les groupes soient connus, ce qui est généralement là où l'affirmation échoue. Le handicap n'est tout simplement pas mesurable de cette façon : les employeurs ne savent pas, et n'ont pas le droit de demander avant une offre conditionnelle, quels candidats sont en situation de handicap, de sorte qu'aucun instrument de ratio d'im-

pact ne peut détecter une discrimination fondée sur le handicap. Une organisation affirmant une atténuation des biais dans la planification de la relève a généralement mesuré la race et le sexe, si elle a mesuré quoi que ce soit, et n'a rien dit sur la population la plus affectée par les modèles pilotés par les données de performance.

La planification de la relève comporte une difficulté supplémentaire. Un modèle identifiant la préparation ou le potentiel à partir de l'ancienneté, du stade de carrière, du temps passé dans le poste, de la fraîcheur des diplômes et du rythme des avancements précédents s'appuie sur des signaux fortement corrélés à l'âge. Rapporter une atténuation des biais pour un tel système, sans avoir examiné sa répartition par âge, affirme l'inverse de ce que prédisent les données d'entrée.

Et l'affirmation elle-même crée une exposition. Le Titre VII, tel que modifié en 1991, rend illégal, dans le cadre de la sélection ou de la recommandation de candidats pour un emploi ou une promotion, d'ajuster les scores, d'utiliser des seuils différents, ou d'altérer de toute autre manière les résultats de tests liés à l'emploi en fonction de la race, de la couleur, de la religion, du sexe ou de l'origine nationale. Rien ici ne suggère que ces systèmes font cela. Le point est qu'une déclaration publique reliant un outil automatisé de promotion à des résultats démographiques est un document que l'organisation a rédigé, ne peut pas retirer, et pourrait être amenée à expliquer.

Ce qui manque dans les quatre cas

Chacun des échecs précédents est une version de la même absence. Aucune de ces études de cas ne contient de contrefactuel, c'est-à-dire une description de ce qui se serait passé sans l'intervention.

L'organisation a déployé un système pendant une période où le marché du travail a évolué, les effectifs ont changé, les managers se sont renouvelés, la rémunération a été ajustée, et d'autres initiatives ont été menées en parallèle. Elle a ensuite attribué le mouvement de quatre mesures à l'un de ces changements. Ce n'est pas une petite objection méthodologique. C'est la question entière, et son omission est ce qui rend le genre infalsifiable.

Les preuves publiées sur des affirmations connexes devraient donner matière à réflexion à tout acheteur. Lorsque la Harvard Business School et le Burning Glass Institute ont examiné ce qui s'était réellement passé dans les entreprises ayant publiquement adopté le recrutement fondé sur les compétences, ils ont constaté que l'abandon

des exigences de diplôme avait fait évoluer l'embauche de candidats non diplômés d'environ trois points et demi de pourcentage, soit moins d'une embauche sur sept cents, près de la moitié des entreprises ne montrant aucun changement de comportement malgré leurs annonces. Les annonces étaient sincères. La pratique a complètement divergé, et les organisations qui les faisaient ne le savaient pas.

L'inférence pertinente n'est pas que les gens mentent. C'est que les organisations, de manière fiable, ne savent pas si leurs propres initiatives ont fonctionné, et qu'une étude de cas est rédigée par la partie la moins bien placée pour le découvrir.

Pourquoi le demandeur y croit généralement

Il vaut la peine de le dire directement, car la lecture alternative est malveillante et généralement erronée.

Ces chiffres sont rarement fabriqués. Ils sont produits par un environnement de mesure dans lequel aucune référence n'a été établie, les définitions ont changé, le fournisseur a fourni le cadre analytique, et la personne qui compile les résultats est celle qui a défendu l'achat. Chacune de ces conditions biaise le résultat dans le même sens, et aucune ne requiert que quiconque agisse de mauvaise foi. Le chiffre est réel en ce sens que quelqu'un l'a calculé. Il ne signifie simplement pas ce que la phrase qui l'entoure prétend.

C'est aussi pourquoi l'environnement réglementaire a commencé à évoluer. La Federal Trade Commission a été active contre les affirmations non étayées sur les capacités de l'intelligence artificielle, y compris une ordonnance de janvier 2025 obligeant un fournisseur de logiciels d'accessibilité à payer un million de dollars pour avoir affirmé que son produit automatisé pouvait rendre n'importe quel site web conforme aux normes d'accessibilité. Le principe pertinent ne se limite pas à cette catégorie de produits. Les affirmations de performance concernant l'intelligence artificielle requièrent une justification, et une organisation publiant des pourcentages de résultats dans un document marketing fait précisément ce type d'affirmation.

Ce qu'un acheteur devrait demander

Quatre questions, qui prennent environ dix minutes et auxquelles la grande majorité des études de cas ne survivent pas.

Quelle était la référence, quand a-t-elle été mesurée, et par qui. Si la référence a été construite après le déploiement, ou fournie par le fournisseur, la comparaison n'en est pas une.

La définition a-t-elle changé. Pour toute mesure impliquant une population désignée, comme les hauts potentiels ou les postes clés, demandez si les critères de désignation étaient les mêmes avant et après. Si le système les a modifiés, les populations diffèrent et la comparaison est nulle.

Qu'est-ce qui a changé d'autre pendant la période. Marché du travail, effectifs, rémunération, management, initiatives concurrentes. Un répondant honnête aura une liste. Un répondant qui n'en a pas n'a pas cherché.

Quel est le dénominateur, et combien de personnes. Les pourcentages sans taille de cohorte sont ininterprétables, et les petites cohortes produisent de grands pourcentages à partir de petits mouvements.

> **UNE CINQUIÈME QUESTION**
>
> Une cinquième question mérite d'être posée à propos de toute allégation touchant au biais : quels groupes ont été mesurés, à quelle taille de cellule, et qu'a-t-on fait pour les groupes qui ne peuvent pas être mesurés. La réponse à cette dernière partie est presque toujours que rien ne l'a été, ce qui est en soi l'enseignement le plus utile que l'acheteur en tirera.

Ce que contiendrait une étude de cas crédible

Le genre n'est pas irrécupérable, et une organisation disposée à publier ce qui suit se distinguerait immédiatement dans un domaine où personne ne le fait.

Indiquez la référence de base et sa date. Indiquez les définitions et signalez celles qui ont changé. Rapportez la taille des cohortes en même temps que les pourcentages. Nommez les changements concomitants qui pourraient expliquer l'évolution, et précisez ceux que vous ne pouvez pas exclure. Décrivez ce qui n'a pas fonctionné, car un déploiement dans lequel tout s'est amélioré est un déploiement dans lequel rien n'a été mesuré. Et lorsqu'une allégation de biais est formulée, indiquez les groupes, la méthode, la taille des cellules, et ce qui reste non mesuré.

Ce document est moins impressionnant que la forme actuelle et considérablement plus utile. C'est aussi la seule version qui résiste à la lecture d'une personne soucieuse de savoir si elle est vraie.

Conclusion

L'étude de cas est l'artefact probant principal dans l'achat de technologies de ressources humaines en entreprise, et c'est la forme de preuve la plus faible en usage professionnel courant. Elle concerne une seule organisation, n'est pas contrôlée, repose sur une auto-déclaration, est définie rétrospectivement, et est publiée par une partie ayant un intérêt dans le résultat.

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.

Références

- Federal Trade Commission, *In the Matter of accessiBe, Inc.*, dossier n° 2223156, plainte et ordonnance proposée annoncées le 3 janvier 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, février 2024.
- Titre VII du Civil Rights Act de 1964, tel que modifié, 42 U.S.C. § 2000e-2, y compris § 2000e-2(l).
- Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. Part 1607, applicable aux procédures de sélection affectant l'embauche, la promotion et d'autres opportunités d'emploi.

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.

Comment citer

Grizzle, H. M. (2025, May 23). Quatre chiffres et aucun contrefactuel : lecture de l'étude de cas sur l'intelligence artificielle en ressources humaines. Novara Consulting Group. https://www.novaracg.com/fr/2025/05/23/four-numbers-and-no-counterfactual-reading-the-artificial-intelligence-case-study-in-human-resources/