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/.

Quatro Números e Nenhum Contrafactual: Lendo o Estudo de Caso de Inteligência Artificial em Recursos Humanos

Heather Grizzle · May 23, 2025

Quatro melhorias percentuais e nenhum contrafactual. O que cada métrica padrão de estudo de caso de IA em RH exige que seja verdadeiro, e as perguntas que levam dez minutos.

Conteúdo

1. O género
2. Afirmação um: redução no tempo de preenchimento de vagas
3. Afirmação dois: redução da atividade transacional manual
4. Afirmação três: aumento da retenção de altos potenciais
5. Afirmação quatro: mitigação de enviesamento
6. O que falta em todas as quatro
7. Porque é que o autor da afirmação geralmente acredita nela
8. O que um comprador deve perguntar
9. O que um estudo de caso credível deveria conter
10. Conclusão
11. Referências

O estudo de caso de inteligência artificial em recursos humanos consolidou-se numa forma estável.

O género

O estudo de caso de inteligência artificial em recursos humanos consolidou-se numa forma estável. Uma organização descreve uma implementação, reporta três ou quatro pontos percentuais de melhoria, acrescenta uma afirmação sobre mitigação de enviesamento, e conclui que o talento se tornou uma fonte de vantagem competitiva. Os números são específicos. A metodologia está ausente. A conclusão convida a uma conversa.

A forma persiste porque funciona como persuasão e porque ninguém está posicionado para a verificar. Os fornecedores publicam-na porque vende. Os profissionais publicam-na porque estabelece estatuto. Os compradores aceitam-na porque a alternativa é admitir que não têm base para a decisão que estão prestes a tomar. Todos na cadeia têm interesse no número e ninguém tem interesse no denominador.

Todos na cadeia têm interesse no número e ninguém tem interesse no denominador.

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.

Afirmação um: redução no tempo de preenchimento de vagas

O tempo de preenchimento de vagas é a métrica mais comummente reportada e a mais facilmente manipulável sem alterar nada no processo de contratação.

A medição requer um ponto de partida definido e um ponto final definido, mantidos constantes entre os períodos de comparação. Na prática, o ponto de partida varia consoante a organização e frequentemente consoante a requisição: pode ser a data em que a requisição foi aprovada, a data em que foi publicada, a data em que o recrutamento começou, ou a data em que o gestor de contratação levantou pela primeira vez a necessidade. Mover esse ponto de partida para mais tarde encurta o intervalo medi-

do sem encurtar nada que alguém tenha realmente experienciado. O mesmo acontece ao excluir requisições canceladas ou retiradas, ao excluir as funções que demoraram mais tempo, ou ao alterar a composição de funções no denominador.

O fator de confusão é maior do que o efeito. O tempo de preenchimento de vagas responde fortemente às condições do mercado de trabalho, e os períodos de comparação relevantes para a maioria dos estudos de caso atuais abrangem uma mudança de escassez aguda de mão-de-obra para condições consideravelmente mais soltas. Uma organização que não tivesse alterado nada mostraria melhoria ao longo dessa janela temporal. Atribuir essa evolução a um sistema implementado durante esse período exige um contrafactual, e nenhum estudo de caso deste género o contém.

O qualificador também merece atenção. Uma melhoria reportada para funções-chave, em vez de todas as funções, indica um subconjunto, e quando esse subconjunto foi selecionado depois de os resultados serem conhecidos, o valor descreve a seleção e não o sistema.

Afirmação dois: redução da atividade transacional manual

Esta afirmação requer uma medição de referência da atividade manual, e quase nenhuma função de recursos humanos a possui.

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.

Afirmação três: aumento da retenção de altos potenciais

Esta é a mais interessante das quatro, porque o mecanismo que produz o número é habitualmente a própria implementação.

Alto potencial não é uma característica observada. É uma designação que a organização atribui, e um sistema de talento com inteligência artificial tipicamente altera a forma como essa designação é feita. Esse é o propósito declarado. Se o sistema altera quais os colaboradores designados como de alto potencial, então a coorte medida após a implementação é uma população diferente da coorte medida antes, e uma comparação de retenção entre elas não é uma comparação da mesma coisa.

A direção do enviesamento resultante é previsível. Um modelo que identifica alto potencial a partir do historial de desempenho, sinais de envolvimento e trajetória de carreira irá designar preferencialmente colaboradores que estão envolvidos, recentemente promovidos e em progressão, tudo isto entre os melhores preditores disponíveis de permanência. O sistema seleciona pessoas que já iriam permanecer e a organização regista essa permanência como um efeito do sistema.

A dimensão da coorte agrava isto. Na maioria das organizações, a população de alto potencial é suficientemente pequena para que uma variação de dezoito por cento represente um punhado de indivíduos, bem dentro da variação que qualquer coorte pequena produz de ano para ano. Um valor deste tipo deveria ser acompanhado da dimensão da coorte e de uma indicação de se a definição mudou. Quase nunca é.

Afirmação quatro: mitigação de enviesamento

A quarta afirmação é diferente das outras em natureza, porque não é meramente infundada. É uma representação legal.

Para a fundamentar é necessário medir o enviesamento antes e depois, em grupos definidos, com dimensões de célula adequadas para sustentar a comparação. Requer também que os grupos sejam conhecidos, o que é onde a afirmação geralmente falha. A deficiência não é mensurável desta forma de todo: os empregadores não sabem e podem não poder legalmente perguntar, antes de uma oferta condicional, quais os candidatos com deficiência, pelo que nenhum instrumento de rácio de impacto pode registar discriminação por deficiência. Uma organização que afirma mitigação de enviesamento no planeamento de sucessão geralmente mediu raça e sexo, se é que mediu

alguma coisa, e não disse nada sobre a população mais afetada por modelos orientados por dados de desempenho.

O planeamento de sucessão traz uma dificuldade adicional. Um modelo que identifica prontidão ou potencial a partir da antiguidade, fase da carreira, tempo na função, recência das credenciais e ritmo de progressões anteriores está a basear-se em sinais fortemente correlacionados com a idade. Reportar mitigação de enviesamento para um tal sistema, sem ter examinado a sua distribuição etária, afirma o oposto do que os dados de entrada preveem.

E a própria afirmação cria exposição. O Título VII, conforme alterado em 1991, torna ilegal, em relação à seleção ou encaminhamento de candidatos para emprego ou promoção, ajustar pontuações, usar diferentes pontuações de corte, ou de outra forma alterar os resultados de testes relacionados com emprego com base em raça, cor, religião, sexo ou origem nacional. Nada aqui sugere que estes sistemas o façam. O ponto é que uma declaração pública que liga uma ferramenta automatizada de promoção a resultados demográficos é um documento que a organização escreveu, não pode retirar, e pode ser chamada a explicar.

O que falta em todas as quatro

Cada uma das falhas anteriores é uma versão da mesma ausência. Nenhum destes estudos de caso contém um contrafactual, ou seja, uma descrição do que teria acontecido sem a intervenção.

A organização implementou um sistema durante um período em que o mercado de trabalho mudou, o número de colaboradores variou, os gestores mudaram, a remuneração foi ajustada, e outras iniciativas decorreram em simultâneo. Depois atribuiu a variação em quatro métricas a uma dessas mudanças. Isto não é uma pequena questiúncula metodológica. É a questão inteira, e a sua omissão é o que torna o género infalsificável.

A evidência publicada sobre afirmações adjacentes deveria dar que pensar a qualquer comprador. Quando a Harvard Business School e o Burning Glass Institute examinaram o que realmente aconteceu em empresas que adotaram publicamente a contratação baseada em competências, descobriram que eliminar os requisitos de grau académico moveu a contratação de candidatos sem grau académico em cerca de três vírgula cinco pontos percentuais, correspondendo a menos de uma em setecentas contrata-

ções, com quase metade das empresas a não mostrar qualquer mudança comportamental apesar dos seus anúncios. Os anúncios eram sinceros. A prática divergiu completamente, e as organizações que os fizeram não sabiam disso.

A inferência relevante não é que as pessoas estejam a mentir. É que as organizações tipicamente não sabem se as suas próprias iniciativas funcionaram, e que um estudo de caso é escrito pela parte menos posicionada para o descobrir.

Porque é que o autor da afirmação geralmente acredita nela

Vale a pena afirmar isto diretamente, porque a leitura alternativa é pouco generosa e geralmente errada.

Estes valores raramente são fabricados. São produzidos por um ambiente de medição em que nenhuma referência foi estabelecida, as definições mudaram, o fornecedor forneceu o quadro analítico, e a pessoa que compilou os resultados é a pessoa que defendeu a compra. Cada uma destas condições enviesa o resultado na mesma direção, e nenhuma delas exige que alguém aja de má-fé. O número é real no sentido em que alguém o calculou. Simplesmente não significa o que a frase à sua volta diz.

É também por isto que o ambiente regulatório começou a mover-se. A Federal Trade Commission tem estado ativa contra afirmações não sustentadas sobre capacidades de inteligência artificial, incluindo uma decisão de janeiro de 2025 que exigiu a um fornecedor de software de acessibilidade o pagamento de um milhão de dólares por representações de que o seu produto automatizado poderia trazer qualquer sítio web à conformidade com normas de acessibilidade. O princípio relevante não se limita a essa categoria de produto. As afirmações de desempenho sobre inteligência artificial requerem fundamentação, e uma organização que publica percentagens de resultados num documento de marketing está a fazer uma afirmação exatamente desse tipo.

O que um comprador deve perguntar

Quatro perguntas, que levam cerca de dez minutos e às quais a grande maioria dos estudos de caso não resiste.

Qual era a referência de base, quando foi medida, e por quem. Se a referência de base foi construída depois da implementação, ou fornecida pelo fornecedor, a comparação não é uma comparação.

A definição mudou. Para qualquer métrica que envolva uma população designada, como alto potencial ou funções-chave, pergunte se os critérios de designação eram os mesmos antes e depois. Se o sistema os alterou, as populações diferem e a comparação é nula.

O que mais mudou durante o período. Mercado de trabalho, número de colaboradores, remuneração, gestão, iniciativas concorrentes. Um respondente honesto terá uma lista. Um respondente sem uma não analisou a questão.

Qual é o denominador, e quantas pessoas. Percentagens sem dimensões de coorte são ininterpretáveis, e coortes pequenas produzem grandes percentagens a partir de pequenas variações.

> **UMA QUINTA PERGUNTA**
>
> Uma quinta pergunta vale a pena fazer sobre qualquer alegação que toque no viés: que grupos foram medidos, em que tamanhos de amostra, e o que foi feito em relação aos grupos que não podem ser medidos. A resposta à última parte é quase sempre que nada foi feito, o que é em si a informação mais útil que o comprador irá aprender.

O que um estudo de caso credível deveria conter

O género não está além de reparo, e uma organização disposta a publicar o seguinte distinguir-se-ia imediatamente num campo onde ninguém o faz.

Declare a linha de base e a sua data. Declare as definições e assinale quaisquer que tenham mudado. Reporte os tamanhos das coortes juntamente com as percentagens. Nomeie as mudanças concorrentes que poderiam explicar a variação, e diga quais não pode excluir. Descreva o que não funcionou, porque uma implementação em que tudo melhorou é uma implementação em que nada foi medido. E onde é feita uma alegação de viés, declare os grupos, o método, os tamanhos de amostra, e o que permanece por medir.

Esse documento é menos impressionante do que a forma atual e consideravelmente mais útil. É também a única versão que sobrevive à leitura por parte de alguém com interesse em saber se é verdadeira.

Conclusão

O estudo de caso é o artefacto probatório primário na aquisição de tecnologia de recursos humanos empresarial, e é a forma mais fraca de evidência em uso profissional habitual. É de uma única organização, não controlado, autorrelatado, definido retrospetivamente, e publicado por uma parte com interesse na conclusão.

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.

Referências

- Federal Trade Commission, *In the Matter of accessiBe, Inc.*, Processo n.º 2223156, queixa e proposta de decisão anunciadas em 3 de janeiro de 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, fevereiro de 2024.
- Título VII da Lei dos Direitos Civis de 1964, conforme alterada, 42 U.S.C. § 2000e-2, incluindo o § 2000e-2(l).
- Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. Parte 1607, aplicáveis a procedimentos de seleção que afetam a contratação, promoção, e outras oportunidades de emprego.

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.

Como citar

Grizzle, H. M. (2025, May 23). Quatro Números e Nenhum Contrafactual: Lendo o Estudo de Caso de Inteligência Artificial em Recursos Humanos. Novara Consulting Group. https://www.novaracg.com/pt/2025/05/23/four-numbers-and-no-counterfactual-reading-the-artificial-intelligence-case-study-in-human-resources/