NOVARA
CONSULTING GROUP

Este documento fue traducido del inglés por una máquina y puede contener errores. El original en inglés está disponible en https://www.novaracg.com/2025/05/23/four-numbers-and-no-counterfactual-reading-the-artificial-intelligence-case-study-in-human-resources/.

# Cuatro números y ninguna hipótesis contrafactual: lectura del estudio de caso sobre inteligencia artificial en Recursos Humanos

Heather Grizzle · 23 de mayo de 2025

*Cuatro mejoras porcentuales y ningún contrafactual. Lo que cada métrica estándar de un estudio de caso de IA en RR. HH. exige que sea cierto, y las preguntas que llevan diez minutos.*

**Contenidos**

1. El género
2. Afirmación uno: reducción del tiempo de contratación
3. Afirmación dos: reducción de la actividad transaccional manual
4. Afirmación tres: aumento en la retención de alto potencial
5. Afirmación cuatro: mitigación de sesgos
6. Lo que falta en las cuatro
7. Por qué el demandante suele creerlo
8. Lo que un comprador debería preguntar
9. Qué contendría un caso de estudio creíble
10. Conclusión
11. Referencias

*El caso práctico de la inteligencia artificial en recursos humanos se ha asentado en una forma estable.*

## El género

El caso práctico de la inteligencia artificial en recursos humanos se ha asentado en una forma estable. Una organización describe una implementación, informa mejoras de tres o cuatro puntos porcentuales, añade una afirmación sobre la mitigación de sesgos y concluye que el talento se ha convertido en una fuente de ventaja competitiva. Las cifras son específicas. La metodología está ausente. La conclusión invita a una conversación.

La forma persiste porque funciona como persuasión y porque nadie está en posición de verificarla. Los proveedores la publican porque vende. Los profesionales la publican porque establece reputación. Los compradores la aceptan porque la alternativa es admitir que no tienen base para la decisión que están a punto de tomar. Todos en la cadena tienen interés en la cifra y nadie tiene interés en el denominador.

Todos en la cadena tienen interés en la cifra y nadie tiene interés en el denominador.

La posición de Novara Consulting Group es que estos documentos deben leerse como afirmaciones publicitarias más que como evidencia, que las cifras específicas más citadas son las que se producen más fácilmente sin ninguna mejora subyacente, y que un comprador capaz de hacer cuatro preguntas puede distinguir un resultado real de uno reportado en unos diez minutos. Lo que sigue es ese método, aplicado a las cuatro afirmaciones que aparecen con mayor frecuencia.

## Afirmación uno: reducción del tiempo de contratación

El tiempo de contratación es la métrica más comúnmente reportada y la más fácil de mover sin cambiar nada respecto a la contratación.

La medición requiere un punto de inicio definido y un punto final definido, mantenidos constantes a lo largo de los períodos de comparación. En la práctica, el punto de inicio varía según la organización y a menudo según la solicitud: puede ser la fecha en que se aprobó la solicitud, la fecha en que se publicó, la fecha en que comenzó la búsqueda de candidatos, o la fecha en que el gerente de contratación planteó por primera vez la necesidad. Mover ese punto de inicio más tarde acorta el intervalo medido sin acortar nada experimentado por nadie. Lo mismo ocurre al excluir solicitudes cance-

ladas o retiradas, excluir los puestos que tardaron más, o cambiar la composición de los puestos en el denominador.

La variable de confusión es mayor que el efecto. El tiempo de contratación responde fuertemente a las condiciones del mercado laboral, y los períodos de comparación relevantes para la mayoría de los casos prácticos actuales abarcan un cambio de escasez aguda de mano de obra a condiciones considerablemente más laxas. Una organización que no hubiera cambiado nada mostraría mejoras en ese lapso. Atribuir el movimiento a un sistema implementado durante ese período requiere un contrafactual, y ningún caso práctico de este género lo contiene.

El calificativo también merece atención. Una mejora reportada para puestos clave en lugar de todos los puestos indica un subconjunto, y cuando el subconjunto se seleccionó después de conocer los resultados, la cifra describe la selección y no el sistema.

## Afirmación dos: reducción de la actividad transaccional manual

Esta afirmación requiere una medición de referencia de la actividad manual, y casi ninguna función de recursos humanos la tiene.

Cuando existe una cifra, generalmente se ha obtenido de una de tres maneras: a partir del modelo de implementación de un proveedor, de una encuesta en la que el personal estimó su propia distribución del tiempo, o de un recuento de tickets o transacciones cuya definición cambió cuando se introdujo el nuevo sistema. Ninguna de estas opciones respalda una comparación porcentual a lo largo del tiempo, porque el instrumento cambió en el mismo momento que la intervención.

También existe una cuestión de alcance que la métrica oculta. El trabajo transaccional que desaparece de la función de recursos humanos con frecuencia reaparece en otro lugar, más comúnmente entre empleados y gerentes de línea a través del autoservicio. Una reducción del sesenta por ciento en las transacciones gestionadas por la función puede representar una ganancia genuina de eficiencia, una transferencia del mismo trabajo a personas cuyo tiempo no se está contabilizando, o alguna combinación de ambas. Distinguir esto requiere medir el esfuerzo organizacional total, algo que nadie hace, porque la métrica existe para justificar la inversión de la función en lugar de describir el costo de la organización.

## Afirmación tres: aumento en la retención de alto potencial

Esta es la más interesante de las cuatro, porque el mecanismo que produce la cifra suele ser la propia implementación.

Alto potencial no es una característica observada. Es una designación que la organización asigna, y un sistema de talento con inteligencia artificial típicamente cambia cómo se asigna. Ese es el propósito declarado. Si el sistema altera qué empleados son designados como de alto potencial, entonces la cohorte medida después de la implementación es una población distinta de la cohorte medida antes, y una comparación de retención entre ambas no es una comparación de lo mismo.

La dirección del sesgo resultante es predecible. Un modelo que identifica alto potencial a partir del historial de desempeño, señales de compromiso y trayectoria profesional designará preferentemente a empleados que están comprometidos, recientemente ascendidos y en progreso, todos ellos entre los predictores disponibles más sólidos de permanencia. El sistema selecciona a personas que de todos modos iban a quedarse, y la organización registra esa permanencia como un efecto del sistema.

El tamaño de la cohorte agrava esto. En la mayoría de las organizaciones, la población de alto potencial es lo suficientemente pequeña como para que un cambio del dieciocho por ciento represente un puñado de individuos, muy dentro de la variación que cualquier cohorte pequeña produce año tras año. Una cifra de este tipo debería ir acompañada del tamaño de la cohorte y de si la definición cambió. Casi nunca es así.

## Afirmación cuatro: mitigación de sesgos

La cuarta afirmación es diferente de las demás en su naturaleza, porque no es simplemente infundada. Es una declaración legal.

Sustentarla requiere medir el sesgo antes y después, en grupos definidos, con tamaños de celda adecuados para respaldar la comparación. También requiere que los grupos sean conocidos, que es donde la afirmación suele fallar. La discapacidad no es medible de esta manera en absoluto: los empleadores no saben y legalmente no pueden preguntar antes de una oferta condicional qué solicitantes tienen discapacidades, por lo que ningún instrumento de razón de impacto puede registrar la discriminación por discapacidad. Una organización que afirma mitigar el sesgo en la planificación de sucesión generalmente ha medido raza y sexo, si es que ha medido algo, y no ha dicho

nada sobre la población más afectada por los modelos impulsados por datos de desempeño.

La planificación de sucesión presenta una dificultad adicional. Un modelo que identifica preparación o potencial a partir de la antigüedad, la etapa profesional, el tiempo en el puesto, la vigencia de las credenciales y el ritmo de ascensos previos se basa en señales altamente correlacionadas con la edad. Informar mitigación de sesgos para un sistema así, sin haber examinado su distribución por edad, afirma lo contrario de lo que predicen los datos de entrada.

Y la afirmación misma crea exposición. El Título VII, tal como fue enmendado en 1991, hace ilegal, en relación con la selección o remisión de candidatos para empleo o ascenso, ajustar puntuaciones, usar diferentes puntuaciones de corte, o alterar de otro modo los resultados de pruebas relacionadas con el empleo por motivos de raza, color, religión, sexo u origen nacional. Nada aquí sugiere que estos sistemas hagan eso. El punto es que una declaración pública que vincula una herramienta automatizada de ascensos con resultados demográficos es un documento que la organización escribió, no puede retirar, y sobre el cual puede que se le pida explicar.

## Lo que falta en las cuatro

Cada uno de los fallos anteriores es una versión de la misma ausencia. Ninguno de estos casos prácticos contiene un contrafactual, es decir, una descripción de lo que habría sucedido sin la intervención.

La organización implementó un sistema durante un período en el que el mercado laboral cambió, la plantilla varió, los gerentes rotaron, la compensación se ajustó, y otras iniciativas se llevaron a cabo simultáneamente. Luego atribuyó el movimiento en cuatro métricas a uno de esos cambios. Esto no es un pequeño reparo metodológico. Es la cuestión entera, y su omisión es lo que hace que el género sea infalsable.

La evidencia publicada sobre afirmaciones adyacentes debería hacer que cualquier comprador se detenga a pensar. Cuando Harvard Business School y el Burning Glass Institute examinaron lo que realmente sucedió en empresas que públicamente adoptaron la contratación basada en habilidades, descubrieron que eliminar los requisitos de titulación movió la contratación de candidatos sin título en aproximadamente tres punto cinco puntos porcentuales, lo que equivale a menos de una de cada setecientas contrataciones, con casi la mitad de las empresas sin mostrar ningún cambio de com-

portamiento pese a sus anuncios. Los anuncios eran sinceros. La práctica divergió por completo, y las organizaciones que los hicieron no lo sabían.

La inferencia relevante no es que la gente esté mintiendo. Es que las organizaciones sistemáticamente no saben si sus propias iniciativas funcionaron, y que un caso práctico lo escribe la parte menos capacitada para averiguarlo.

## Por qué el demandante suele creerlo

Vale la pena decir esto directamente, porque la lectura alternativa es poco caritativa y generalmente errónea.

Estas cifras rara vez son fabricadas. Se producen en un entorno de medición en el que no se estableció ninguna línea base, las definiciones cambiaron, el proveedor suministró el marco analítico, y la persona que compila los resultados es la misma que abogó por la compra. Cada una de esas condiciones sesga el resultado en la misma dirección, y ninguna de ellas requiere que alguien actúe de mala fe. La cifra es real en el sentido de que alguien la calculó. Simplemente no significa lo que dice la frase que la rodea.

Esto también explica por qué el entorno regulatorio ha comenzado a moverse. La Federal Trade Commission ha estado activa contra afirmaciones infundadas sobre capacidades de inteligencia artificial, incluyendo una orden de enero de 2025 que exige a un proveedor de software de accesibilidad pagar un millón de dólares por declaraciones de que su producto automatizado podía hacer que cualquier sitio web cumpliera con los estándares de accesibilidad. El principio relevante no se limita a esa categoría de producto. Las afirmaciones de rendimiento sobre inteligencia artificial requieren sustentación, y una organización que publica porcentajes de resultados en un documento de marketing está haciendo precisamente ese tipo de afirmación.

## Lo que un comprador debería preguntar

Cuatro preguntas, que toman unos diez minutos y que la gran mayoría de los casos prácticos no pueden resistir.

Cuál fue la línea base, cuándo se midió y por quién. Si la línea base se construyó después de la implementación, o la suministró el proveedor, la comparación no es tal.

Si cambió la definición. Para cualquier métrica que involucre una población designada, como alto potencial o puestos clave, pregunte si los criterios de designación fueron





los mismos antes y después. Si el sistema los cambió, las poblaciones difieren y la comparación es nula.

Qué más cambió durante el período. Mercado laboral, plantilla, compensación, gestión, iniciativas concurrentes. Un encuestado honesto tendrá una lista. Un encuestado sin una no ha investigado.

Cuál es el denominador y cuántas personas. Los porcentajes sin tamaños de cohorte son ininterpretables, y las cohortes pequeñas producen porcentajes grandes a partir de movimientos pequeños.

> **UNA QUINTA PREGUNTA**
>
> Una quinta pregunta merece hacerse ante cualquier afirmación que toque el sesgo: qué grupos se midieron, con qué tamaños de celda, y qué se hizo respecto de los grupos que no se pueden medir. La respuesta a esta última parte es casi siempre que no se hizo nada, lo cual es en sí mismo lo más útil que el comprador aprenderá.

## Qué contendría un caso de estudio creíble

El género no está más allá de toda reparación, y una organización dispuesta a publicar lo siguiente se distinguiría de inmediato en un campo en el que nadie lo hace.

Indique la línea de base y su fecha. Indique las definiciones y señale cualquiera que haya cambiado. Reporte los tamaños de cohorte junto con los porcentajes. Nombre los cambios concurrentes que podrían explicar el movimiento, y diga cuáles no puede descartar. Describa lo que no funcionó, porque una implementación en la que todo mejoró es una implementación en la que no se midió nada. Y cuando se haga una afirmación sobre sesgo, indique los grupos, el método, los tamaños de celda y lo que permanece sin medir.

Ese documento es menos impresionante que el formulario actual y considerablemente más útil. Es también la única versión que sobrevive a la lectura de alguien interesado en si es verdadera.

## Conclusión

El caso de estudio es el artefacto probatorio principal en la compra de tecnología de recursos humanos empresarial, y es la forma más débil de evidencia en el uso profesional habitual. Es de una sola organización, no controlado, autoinformado, definido retrospectivamente y publicado por una parte con interés en el hallazgo.

Nada de eso lo hace inútil. Lo convierte en una hipótesis. Las organizaciones que realmente obtendrán retornos de estos sistemas son aquellas que tratan los números de otros como una razón para investigar en lugar de una razón para comprar, y que someten sus propios números al estándar que quisieran aplicar a los de un proveedor. La disciplina no es difícil ni costosa. Consiste en una línea base, una definición estable, un tamaño de cohorte y un relato honesto de qué más estaba ocurriendo en ese momento.

## Referencias


- Comisión Federal de Comercio, *In the Matter of accessiBe, Inc.*, Expediente N.º 2223156, denuncia y orden propuesta anunciadas el 3 de enero de 2025.
- Fuller, Joseph, et al. *La contratación basada en habilidades: el largo camino de las declaraciones a la práctica*. Harvard Business School Project on Managing the Future of Work y The Burning Glass Institute, febrero de 2024.
- Título VII de la Ley de Derechos Civiles de 1964, en su versión enmendada, 42 U.S.C. § 2000e-2, incluyendo el § 2000e-2(l).
- Directrices Uniformes sobre Procedimientos de Selección de Empleados, 29 C.F.R. Parte 1607, aplicables a los procedimientos de selección que afectan la contratación, la promoción y otras oportunidades de empleo.

**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.

---

## Cómo citar

Grizzle, H. M. (2025, May 23). Cuatro números y ninguna hipótesis contrafactual: lectura del estudio de caso sobre inteligencia artificial en Recursos Humanos. Novara Consulting Group. https://www.novaracg.com/es/2025/05/23/four-numbers-and-no-counterfactual-reading-the-artificial-intelligence-case-study-in-human-resources/