

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/>.

El riesgo no valorado no es rentabilidad: por qué la mayoría de los casos de negocio de inteligencia artificial en Recursos Humanos fracasan como casos de negocio

Heather Grizzle · April 4, 2025

Los ahorros que son medibles en la IA de RR. HH. son pequeños. Los ahorros que son grandes no se pueden medir. Lo que incluye un caso de negocio de inteligencia artificial defendible.

Contenidos

1. Las afirmaciones y las mediciones que se necesitarían para respaldarlas
2. Lo que realmente se puede medir, y cuánto vale
3. La calidad de contratación no es una métrica
4. La predicción de la rotación y la brecha de precisión
5. El cumplimiento automatizado no descarga la obligación de cumplimiento
6. Los costos que el modelo omite
7. Una estructura defendible
8. Conclusión
9. Referencias



Las afirmaciones y las mediciones que se necesitarían para respaldarlas

El argumento comercial estándar para la inteligencia artificial en recursos humanos se ha estabilizado en torno a cinco afirmaciones: contratación más rápida y mejor, menor costo administrativo, planificación predictiva de la fuerza laboral, mejor retención y cumplimiento normativo automatizado. Estas afirmaciones se repiten con suficiente consistencia en los materiales de los proveedores, la programación de conferencias y los comentarios profesionales que ahora funcionan como una suposición de fondo en lugar de como proposiciones que requieren evidencia.

Requieren evidencia, y el obstáculo para producirla no es la tecnología. Recursos humanos entró en este período con la infraestructura de medición más débil de cualquier función corporativa importante. Finanzas cuenta con estados auditados. Operaciones tiene tasas de rendimiento y de defectos. Recursos humanos cuenta con un conjunto de métricas que están definidas de manera variable, rara vez validadas, pocas veces establecidas como línea base antes de una intervención y con frecuencia producidas por el mismo proveedor cuyo producto se está evaluando. Una organización que no pudo medir el desempeño de su proceso de contratación antes de implementar una herramienta automatizada no puede medir la mejora después, y lo que reporta como retorno es generalmente la diferencia entre dos números que nunca fueron comparables.

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.

Lo que realmente se puede medir, y cuánto vale

Las cinco afirmaciones estándar se dividen claramente en dos categorías, y esa separación es más instructiva que las propias afirmaciones.

Es un documento de promoción con un símbolo de moneda incluido.



La primera categoría contiene efectos que son genuinamente medibles y comparativamente modestos. El tiempo de ciclo en la contratación es medible, tiene una línea base establecida en la mayoría de los sistemas de seguimiento de candidatos, y sí responde a la automatización de la programación, el filtrado y la correspondencia. El costo administrativo por transacción en nómina y administración de beneficios es medible, y también lo son las tasas de error en esos procesos. Las organizaciones que automatizan estas funciones generalmente logran ahorros, y esos ahorros pueden sustentarse.

Dos matices importan. El primero es que la mayoría de los ahorros documentados en nómina y administración de beneficios provienen de la automatización de flujos de trabajo y del procesamiento basado en reglas, tecnologías que preceden por décadas a la generación actual de inteligencia artificial y que con frecuencia se reetiquetan como IA en el posicionamiento de los proveedores y en los argumentos comerciales contruidos sobre esa base. Atribuir ahorros de una automatización de procesos ya establecida a una compra reciente de IA infla el retorno medido de esa compra y, de manera más consecuente, atribuye erróneamente el mecanismo causal, lo que significa que la organización aprende la lección equivocada sobre qué produjo el beneficio. El segundo matiz es que estos ahorros son, en términos absolutos, poco espectaculares. No justifican por sí solos los compromisos empresariales que se están asumiendo ahora.

La segunda categoría contiene efectos que serían sustanciales si fueran reales y que actualmente no se pueden medir: mejor calidad de contratación, mejor retención y mejor planificación de la fuerza laboral. Estas son las afirmaciones que sostienen el argumento comercial, y lo sostienen precisamente porque su magnitud no está restringida por la medición.

La calidad de contratación no es una métrica

La calidad de contratación es la afirmación que carga el peso en casi todos los argumentos comerciales de tecnología de reclutamiento, y no cuenta con una definición común validada. Según la organización, se operacionaliza como calificación de desempeño en el primer año, retención a noventa días, retención a doce meses, satisfacción del gerente de contratación, tiempo hasta la productividad, o un compuesto de varios de estos, cada uno con propiedades de medición diferentes y ninguno con validez de constructo establecida como medida de la eficacia de la contratación.

La consecuencia es que cualquier mejora afirmada en la calidad de contratación es en gran medida un artefacto de la elección definitoria. Una organización que define la ca-



lidad de contratación como retención a noventa días y despliega una herramienta que filtra candidatos parecidos a empleados anteriores de larga permanencia observará una mejora. No ha aprendido que la contratación mejoró. Ha aprendido que un sistema optimizado para un indicador sustituto movió ese indicador sustituto. Cuando el proveedor suministra tanto la herramienta como el marco de medición, algo común, el hallazgo se acerca más a una tautología que a una evidencia.

Esto importa mucho más allá de la higiene metodológica. Según las Directrices Uniformes sobre Procedimientos de Selección de Empleados (Uniform Guidelines on Employee Selection Procedures) en 29 C.F.R. Parte 1607, se espera que un empleador cuyo proceso de selección muestre un impacto adverso produzca evidencia de que el procedimiento es válido para el puesto. La validación relacionada con el criterio requiere una medida defendible del desempeño laboral. Una organización que no puede decir con precisión qué significa la calidad de contratación para sus propios puestos, y que no puede demostrar que su medida de criterio es en sí misma imparcial, no cuenta con la evidencia que el marco contempla. La debilidad de medición que infla el argumento comercial es la misma debilidad que socava la defensa legal, lo cual es un caso inusual en el que un problema metodológico y un problema de responsabilidad legal son el mismo problema.

La predicción de la rotación y la brecha de precisión

La planificación predictiva de la fuerza laboral merece un tratamiento aparte, porque aquí es donde la distancia entre el rendimiento publicado y el rendimiento real en el despliegue es más amplia y la exposición legal es menos comprendida.

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.

La brecha entre el ochenta por ciento y cifras superiores al noventa por ciento no es una diferencia de redondeo en esta aplicación, porque la tasa base es baja. La rotación voluntaria en la mayoría de las organizaciones está muy por debajo del veinte por ciento anual, lo que significa que un modelo con una precisión general aparentemente fuerte puede seguir produciendo falsos positivos que superan sustancialmente a los verdaderos positivos. En un modelo de rotación, un falso positivo es un empleado identificado incorrectamente como propenso a irse.

La cuestión de gobernanza surge de inmediato, y el enfoque de retorno de inversión la oscurece. ¿Qué hace la organización con ese nombre? Si la respuesta es nada, el modelo no ha producido ningún retorno. Si la respuesta es cualquier cosa, ya sean ofertas diferenciales de retención, inversión alterada en desarrollo, exclusión de asignaciones a largo plazo o consideración cambiada de sucesión, entonces la organización ha tomado una decisión de empleo diferencial sobre un individuo con base en una predicción. Eso es una acción de empleo, y su legalidad depende de qué impulsó la puntuación. Las guías para construir estos modelos habitualmente recomiendan incluir datos demográficos del empleado, con la edad y la antigüedad mencionadas entre las variables que dan contexto al modelo. Un empleador que introduce la edad en un modelo, usa el resultado para asignar inversión en desarrollo y no puede reconstruir por qué un empleado en particular fue puntuado de esa manera ha creado una exposición bajo la Ley de Discriminación por Edad en el Empleo (Age Discrimination in Employment Act) y la ha registrado en el balance como un beneficio.

UN RIESGO OPERATIVO QUE SE AGRAVA

Los comentaristas también han señalado un riesgo operativo que se agrava y que no figura en ningún argumento comercial: las puntuaciones de riesgo de fuga comunicadas a gerentes que no han sido capacitados para actuar sobre ellas pueden por sí mismas producir la salida que el modelo predijo.



El cumplimiento automatizado no descarga la obligación de cumplimiento

La afirmación de que la inteligencia artificial garantiza el cumplimiento de la legislación laboral y de las políticas internas debería eliminarse por completo de los argumentos comerciales, y no porque las herramientas sean inútiles. Los sistemas de monitoreo que señalan descansos omitidos, umbrales de horas extra, anomalías de clasificación o errores en la administración de licencias pueden ser genuinamente valiosos, y el argumento operativo a su favor es razonable.

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.

Los sucesos de enero de 2025 empeoran esto en lugar de mejorarlo. Cuando la EEOC retiró de su sitio web su orientación sobre inteligencia artificial, incluida la asistencia técnica de mayo de 2022 sobre la Ley de Estadounidenses con Discapacidades (Americans with Disabilities Act) y la asistencia técnica de mayo de 2023 sobre el Título VII, los empleadores perdieron el material interpretativo que les indicaba qué prácticas consideraba problemáticas la agencia. Los estatutos no cambiaron. Una herramienta de cumplimiento configurada conforme a una orientación que desde entonces ha sido retirada ahora opera sin un estándar de referencia declarado, y el empleador, no el proveedor, carga con esa brecha.

La automatización del cumplimiento, por lo tanto, añade una obligación de documentación y respuesta en lugar de eliminarla. Esa obligación tiene un costo, el costo es real, y pertenece al lado de los costos del modelo.

Los costos que el modelo omite

La mayoría de los argumentos comerciales de IA en recursos humanos contienen un lado de costos compuesto por tarifas de licencia, servicios de implementación y alguna



asignación para capacitación. Ese es el costo de adquisición del software. No es el costo de operar un sistema automatizado de decisiones de empleo, que además incluye, como mínimo, lo siguiente.

Trabajo de validación, es decir, el análisis que establece que lo que puntúa cada sistema corresponde a un requisito documentado del puesto, junto con la revalidación requerida cuando el sistema o el puesto cambian. Operación de la vía de adaptación, es decir, la capacidad permanente de proporcionar una evaluación alternativa a solicitud, con responsables de decisión designados, un plazo de respuesta definido y un proceso interactivo funcional en lugar de un formulario. Capacidad de retención y reproducción de registros, es decir, la capacidad de producir la puntuación, las variables de entrada y la configuración del sistema vigente para cualquier decisión individual durante todo el período de conservación contemplado por 29 C.F.R. Parte 1602. Monitoreo, es decir, pruebas periódicas de los resultados en lugar de depender de un resumen del proveedor. Cooperación del proveedor en procesos de descubrimiento e investigaciones administrativas, lo cual es costoso, está ausente de los términos comerciales estándar a menos que se negocie, y se cobra al empleador cuando se necesita. Y la propia función de gobernanza interna, es decir, la persona identificada responsable de lo anterior, que en la mayoría de las organizaciones actualmente no existe y debe crearse o contratarse.

Junto a estos costos operativos hay un pasivo contingente que prácticamente ningún argumento comercial calcula, y ya no es hipotético. En agosto de 2023 la EEOC llegó a un acuerdo con iTutorGroup por \$365,000 por un software configurado para rechazar a solicitantes por encima de ciertas edades, lo cual es una cifra concreta para una decisión de configuración simple. En julio de 2024 el Distrito Norte de California permitió que avanzaran reclamos contra un proveedor de software de filtrado bajo la teoría de que el proveedor actuó como agente de los empleadores que usaban sus herramientas, lo cual desestabiliza la suposición común de que la responsabilidad recae por completo en el empleador o en el proveedor y puede asignarse limpiamente por contrato. En marzo de 2025 la Unión Estadounidense por las Libertades Civiles (American Civil Liberties Union, ACLU), la ACLU de Colorado, Public Justice y Eisenberg and Baum presentaron una queja ante la División de Derechos Civiles de Colorado y la EEOC en nombre de un empleado Deaf e indígena en relación con una entrevista automatizada por video y una adaptación de subtítulo denegada, un asunto que ambas empresas niegan. Ese reclamo es individualizado, no requiere estadísticas comparativas y es económico de presentar.



El tratamiento correcto no es adivinar una cifra de acuerdo. Es declarar explícitamente la exposición, señalar que los reclamos de práctica sistémica frecuentemente quedan fuera de la cobertura de las pólizas estándar de responsabilidad por prácticas de empleo, y exigir que los controles de gobernanza que reducen la exposición se financien como parte del despliegue en lugar de posponerse.

Una organización que añade estas líneas frecuentemente encontrará que la inversión sigue teniendo sentido. Ese es un buen resultado y no es el punto. El punto es que el número resultante puede defenderse ante una junta directiva, un auditor y eventualmente un tribunal, algo que el número actual no puede hacer.

Una estructura defendible

Lo siguiente produce un argumento comercial que resiste el escrutinio y que puede ser construido por una función existente de recursos humanos y finanzas.

1. Establecer la línea base antes del despliegue, no después: tiempo de ciclo, costo por contratación, costo administrativo por transacción, tasas de error y rotación por segmento, cada uno medido durante al menos cuatro trimestres previos.
2. Separar los ahorros de automatización de los ahorros de IA. Declarar qué beneficios afirmados provienen del procesamiento de flujos de trabajo basado en reglas que estaría disponible sin la compra de IA, y calcular el precio de esos por separado.
3. Definir explícitamente la calidad de contratación para los propios puestos antes de medir cualquier cambio en ella, declarar la medida de criterio y declarar sus limitaciones. Si no se puede definir, no afirmar una mejora en ella.
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. Para cualquier modelo que produzca puntuaciones a nivel individual, documentar qué acción se deriva de la puntuación antes del despliegue, y hacer que esa acción se revise como una decisión de empleo en lugar de como un resultado analítico.
6. Calcular el precio de las obligaciones operativas mencionadas anteriormente como un costo recurrente, no como una implementación única.



7. Declarar el pasivo contingente como una partida revelada junto con los controles que lo reducen, se le asigne o no una cifra en dólares.
8. Exigir evidencia de validación, datos de desempeño en campo y capacidad de reproducción como condiciones contractuales en el momento de la adquisición. Conservar cualquier negativa, la cual es en sí misma información sobre lo que la organización sabe.

Conclusión

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.

Las organizaciones que reportarán un retorno duradero de estos sistemas no son las que los adoptaron primero ni las que más compraron. Son las que midieron antes de implementar, las que pueden decir qué puntúa cada sistema y por qué eso importa para el puesto, y las que trataron el aparato de gobernanza como parte del costo de propiedad y no como un gasto que evitar. Un retorno que depende de un beneficio no medido y un riesgo no cotizado no es retorno. Es un pronóstico, y la diferencia se hace evidente en el momento en que alguien pide ver el trabajo detrás.

Referencias

- American Civil Liberties Union, ACLU de Colorado, Public Justice, y Eisenberg & Baum LLP. Demanda presentada ante la División de Derechos Civiles de Colorado y la Comisión para la Igualdad de Oportunidades en el Empleo de EE. UU., 19 de marzo de 2025.
- EEOC v. iTutorGroup, Inc., No. 1:22-cv-02565 (E.D.N.Y.), decreto de conciliación, agosto de 2023.
- Mobley v. Workday, Inc., No. 3:23-cv-00770 (N.D. Cal.), resolución sobre moción de desestimación, julio de 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 de mayo de 2022, y *Select Issues: Assessing Adverse Impact in Software, Algorithms, and Artificial Intelligence Used in Employment Selection Procedures Under Title VII*, 18 de mayo de 2023. Ambos retirados del sitio web de la agencia el 27 de enero de 2025.
 - Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. Parte 1607; requisitos de mantenimiento de registros, 29 C.F.R. Parte 1602.
-

Cómo citar

Grizzle, H. M. (2025, April 4). El riesgo no valorado no es rentabilidad: por qué la mayoría de los casos de negocio de inteligencia artificial en Recursos Humanos fracasan como casos de negocio. Novara Consulting Group. <https://www.novaracg.com/es/2025/04/04/unpriced-risk-is-not-return-why-most-artificial-intelligence-business-cases-in-human-resources-fail-as-business-cases/>

