NOVARA
CONSULTING GROUP

This document was translated from English by machine and may contain mistakes. The English original is at https://www.novaracg.com/2025/05/23/four-numbers-and-no-counterfactual-reading-the-artificial-intelligence-case-study-in-human-resources/.

Cztery liczby i brak kontrfaktualności: analiza studium przypadku sztucznej inteligencji w zasobach ludzkich

Heather Grizzle · May 23, 2025

Cztery wskaźniki poprawy procentowej i żaden kontrfaktyczny. Co musi być prawdą, aby każdy standardowy wskaźnik w studium przypadku HR AI się potwierdził, oraz pytania, na które odpowiedź zajmuje dziesięć minut.

Spis treści

1. Gatunek
2. Twierdzenie pierwsze: skrócenie czasu obsadzenia stanowiska
3. Twierdzenie drugie: redukcja ręcznych czynności transakcyjnych
4. Twierdzenie trzecie: wzrost retencji pracowników o wysokim potencjale
5. Twierdzenie czwarte: ograniczenie uprzedzeń
6. Czego brakuje we wszystkich czterech przypadkach
7. Dlaczego zgłaszający zwykle w to wierzy
8. O co powinien zapytać nabywca
9. Co zawierałoby wiarygodne studium przypadku
10. Podsumowanie
11. Bibliografia

Studium przypadku sztucznej inteligencji w dziale zasobów ludzkich przybrało już ustaloną formę.

Gatunek

Studium przypadku sztucznej inteligencji w dziale zasobów ludzkich przybrało już ustaloną formę. Organizacja opisuje wdrożenie, raportuje poprawę o trzy lub cztery procent, dodaje twierdzenie o ograniczeniu uprzedzeń i konkluduje, że talent stał się źródłem przewagi konkurencyjnej. Liczby są konkretne. Metodologia jest nieobecna. Konkluzja zaprasza do rozmowy.

Forma ta utrzymuje się, ponieważ działa jako perswazja i ponieważ nikt nie jest w stanie jej zweryfikować. Dostawcy publikują ją, bo się sprzedaje. Praktycy publikują ją, bo buduje pozycję. Nabywcy ją akceptują, bo alternatywą byłoby przyznanie, że nie mają podstaw do decyzji, którą zamierzają podjąć. Każdy ogniwo tego łańcucha ma interes w liczbie, a nikt nie ma interesu w mianowniku.

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.

Każdy ogniwo tego łańcucha ma interes w liczbie, a nikt nie ma interesu w mianowniku.

Twierdzenie pierwsze: skrócenie czasu obsadzenia stanowiska

Czas obsadzenia stanowiska to najczęściej raportowany wskaźnik i najłatwiejszy do przesunięcia bez zmiany czegokolwiek w samym procesie rekrutacji.

Pomiar wymaga zdefiniowanego punktu początkowego i zdefiniowanego punktu końcowego, utrzymywanych stałymi w porównywanych okresach. W praktyce punkt początkowy różni się między organizacjami, a często nawet między poszczególnymi naborami: może to być data zatwierdzenia wakatu, data jego ogłoszenia, data rozpoczęcia poszukiwań kandydatów lub data, w której menedżer zatrudniający po raz pierwszy zgłosił potrzebę. Przesunięcie tego punktu na później skraca mierzony okres, nie skracając niczego, czego ktokolwiek faktycznie doświadczył. To samo dotyczy wyłącze-

nia anulowanych lub wycofanych naborów, wyłączenia stanowisk, których obsadzenie trwało najdłużej, lub zmiany składu stanowisk w mianowniku.

Czynnik zakłócający jest większy niż sam efekt. Czas obsadzenia stanowiska silnie reaguje na warunki rynku pracy, a odpowiednie okresy porównawcze dla większości aktualnych studiów przypadku obejmują przejście od ostrego niedoboru siły roboczej do warunków znacznie luźniejszych. Organizacja, która nie zmieniła niczego, wykazałaby poprawę w tym okresie. Przypisanie tej zmiany systemowi wdrożonemu w tym czasie wymaga kontrfaktycznego porównania, a żadne studium przypadku w tym gatunku go nie zawiera.

Zastrzeżenie również zasługuje na uwagę. Poprawa zaraportowana dla kluczowych stanowisk, a nie dla wszystkich stanowisk, wskazuje na podzbiór, a tam, gdzie podzbiór ten wybrano po poznaniu wyników, liczba ta opisuje selekcję, a nie system.

Twierdzenie drugie: redukcja ręcznych czynności transakcyjnych

To twierdzenie wymaga pomiaru wyjściowego poziomu czynności ręcznych, a niemal żadna funkcja zasobów ludzkich go nie posiada.

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.

Twierdzenie trzecie: wzrost retencji pracowników o wysokim potencjale

Jest to najciekawsze z czterech twierdzeń, ponieważ mechanizm produkujący tę liczbę to zazwyczaj samo wdrożenie.

Wysoki potencjał nie jest cechą obserwowalną. Jest to oznaczenie przypisywane przez organizację, a system talentów oparty na sztucznej inteligencji zazwyczaj zmienia sposób, w jaki jest ono przypisywane. To jest deklarowany cel. Jeśli system zmienia, którzy pracownicy są oznaczani jako osoby o wysokim potencjale, wówczas kohorta mierzona po wdrożeniu stanowi inną populację niż kohorta mierzona przed wdrożeniem, a porównanie retencji między nimi nie jest porównaniem tego samego.

Kierunek wynikającego z tego błędu jest przewidywalny. Model identyfikujący wysoki potencjał na podstawie historii wyników, sygnałów zaangażowania i ścieżki kariery będzie preferencyjnie oznaczał pracowników zaangażowanych, niedawno awansowanych i rozwijających się — a wszystkie te cechy należą do najsilniejszych dostępnych predyktorów pozostania w firmie. System wybiera ludzi, którzy i tak by zostali, a organizacja odnotowuje ich pozostanie jako efekt działania systemu.

Wielkość kohorty tylko to pogłębia. W większości organizacji populacja osób o wysokim potencjale jest na tyle mała, że zmiana o osiemnaście procent reprezentuje garstkę osób, mieszczącą się dobrze w granicach zwykłej zmienności, jaką każda mała kohorta generuje rok do roku. Liczbie tego rodzaju powinna towarzyszyć informacja o wielkości kohorty oraz o tym, czy definicja się zmieniła. Niemal nigdy tak się nie dzieje.

Twierdzenie czwarte: ograniczenie uprzedzeń

Czwarte twierdzenie różni się od pozostałych jakościowo, ponieważ nie jest ono jedynie niepoparte dowodami. Jest to oświadczenie o charakterze prawnym.

Jego udowodnienie wymaga zmierzenia uprzedzeń przed i po, dla zdefiniowanych grup, z wielkościami prób wystarczającymi do przeprowadzenia porównania. Wymaga to także, by grupy te były znane, a to właśnie tu twierdzenie zazwyczaj zawodzi. Niepełnosprawność w ogóle nie jest mierzalna w ten sposób: pracodawcy nie wiedzą i nie mogą zgodnie z prawem pytać przed złożeniem warunkowej oferty, którzy kandydaci są osobami z niepełnosprawnością, więc żaden instrument oparty na wskaźniku wpły-

wu nie może zarejestrować dyskryminacji ze względu na niepełnosprawność. Organizacja twierdząca, że ogranicza uprzedzenia w planowaniu sukcesji, zazwyczaj zmierzyła rasę i płeć, jeśli w ogóle coś zmierzyła, i nic nie powiedziała o populacji najbardziej dotkniętej modelami opartymi na danych o wynikach.

Planowanie sukcesji niesie ze sobą dodatkową trudność. Model identyfikujący gotowość lub potencjał na podstawie stażu pracy, etapu kariery, czasu w danej roli, aktualności kwalifikacji i tempa wcześniejszych awansów opiera się na sygnałach silnie skorelowanych z wiekiem. Raportowanie ograniczenia uprzedzeń dla takiego systemu, bez zbadania jego rozkładu wiekowego, twierdzi coś odwrotnego niż to, co wynika z danych wejściowych.

A samo to twierdzenie stwarza ryzyko prawne. Tytuł VII, w brzmieniu zmienionym w 1991 roku, czyni niezgodnym z prawem, w związku z selekcją lub rekomendowaniem kandydatów do zatrudnienia lub awansu, dostosowywanie wyników, stosowanie różnych progów punktowych lub inne zmienianie wyników testów związanych z zatrudnieniem na podstawie rasy, koloru skóry, religii, płci lub pochodzenia narodowego. Nic tu nie wskazuje, by te systemy tak działały. Chodzi o to, że publiczne oświadczenie łączące zautomatyzowane narzędzie awansowe z wynikami demograficznymi jest dokumentem, który organizacja napisała, nie może wycofać i może zostać poproszona o wyjaśnienie.

Czego brakuje we wszystkich czterech przypadkach

Każdy z powyższych błędów jest wersją tej samej nieobecności. Żadne z tych studiów przypadku nie zawiera kontrfaktu, czyli opisu tego, co wydarzyłoby się bez interwencji.

Organizacja wdrożyła system w okresie, w którym rynek pracy się zmienił, zmieniła się liczba zatrudnionych, zmienili się menedżerowie, dostosowano wynagrodzenia, a równocześnie prowadzono inne inicjatywy. Następnie przypisała zmianę w czterech wskaźnikach jednej z tych zmian. To nie jest drobna uwaga metodologiczna. To jest całe zagadnienie, a jego pominięcie sprawia, że ten gatunek jest niefalsyfikowalny.

Opublikowane dowody dotyczące pokrewnych twierdzeń powinny dać nabywcy do myślenia. Kiedy Harvard Business School i Burning Glass Institute zbadały, co faktycznie wydarzyło się w firmach, które publicznie wdrożyły rekrutację opartą na umiejętnościach, stwierdziły, że rezygnacja z wymogów dotyczących dyplomu przesunęła

zatrudnianie kandydatów bez dyplomu o około trzy i pół punktu procentowego, co stanowi mniej niż jedną na siedemset rekrutacji, przy czym blisko połowa firm nie wykazała żadnej zmiany w praktyce mimo swoich ogłoszeń. Ogłoszenia były szczere. Praktyka całkowicie się rozminęła, a organizacje je składające nie wiedziały o tym.

Istotny wniosek nie jest taki, że ludzie kłamią. Jest taki, że organizacje systematycznie nie wiedzą, czy ich własne inicjatywy zadziałały, a studium przypadku pisze strona najmniej zdolna to ustalić.

Dlaczego zgłaszający zwykle w to wierzy

Warto powiedzieć to wprost, ponieważ alternatywna interpretacja jest nieuczciwa i na ogół błędna.

Te liczby rzadko są sfabrykowane. Powstają w środowisku pomiarowym, w którym nie ustalono żadnego punktu odniesienia, definicje się zmieniały, dostawca dostarczył ramy analityczne, a osoba zestawiająca wyniki jest tą samą osobą, która popierała zakup. Każdy z tych warunków obciąża wynik w tym samym kierunku, a żaden z nich nie wymaga działania w złej wierze. Liczba jest prawdziwa w tym sensie, że ktoś ją obliczył. Po prostu nie oznacza tego, co głosi zdanie wokół niej.

To również dlatego otoczenie regulacyjne zaczęło się zmieniać. Federalna Komisja Handlu (Federal Trade Commission) aktywnie działała przeciwko niepopartym twierdzeniom o możliwościach sztucznej inteligencji, w tym w styczniu 2025 roku wydała nakaz zobowiązujący dostawcę oprogramowania dostępnościowego do zapłaty miliona dolarów za twierdzenia, że jego zautomatyzowany produkt może doprowadzić dowolną stronę internetową do zgodności ze standardami dostępności. Istotna zasada nie ogranicza się do tej kategorii produktów. Twierdzenia dotyczące wydajności sztucznej inteligencji wymagają udokumentowania, a organizacja publikująca procentowe wyniki w dokumencie marketingowym składa twierdzenie dokładnie tego rodzaju.

O co powinien zapytać nabywca

Cztery pytania, które zajmują około dziesięciu minut i których zdecydowana większość studiów przypadku nie przetrwa.

Jaki był punkt odniesienia, kiedy został zmierzony i przez kogo. Jeśli punkt odniesienia został skonstruowany po wdrożeniu lub dostarczony przez dostawcę, porównanie nie jest porównaniem.

Czy zmieniła się definicja. Dla każdego wskaźnika obejmującego wyznaczoną populację, takiego jak wysoki potencjał lub kluczowe stanowiska, zapytaj, czy kryteria wyznaczania były takie same przed i po. Jeśli system je zmienił, populacje się różnią, a porównanie jest nieważne.

Co jeszcze zmieniło się w tym okresie. Rynek pracy, liczba zatrudnionych, wynagrodzenia, zarządzanie, równoległe inicjatywy. Uczciwy respondent będzie miał listę. Respondent bez listy nie sprawdził tego.

Jaki jest mianownik i ile osób. Procenty bez podania wielkości kohorty są niemożliwe do zinterpretowania, a małe kohorty generują duże wartości procentowe z małych zmian.

> **PIĄTE PYTANIE**
>
> Piąte pytanie, które warto zadać w odniesieniu do każdego twierdzenia dotyczącego stronniczości, brzmi: które grupy zostały zmierzone, przy jakich liczebnościach komórek i co zrobiono w sprawie grup, których nie można zmierzyć. Odpowiedź na tę ostatnią część niemal zawsze brzmi, że nic nie zrobiono, co samo w sobie jest najbardziej użyteczną informacją, jaką pozyska kupujący.

Co zawierałoby wiarygodne studium przypadku

Ten gatunek nie jest nie do naprawienia, a organizacja skłonna opublikować poniższe natychmiast wyróżniłaby się w dziedzinie, w której nikt tego nie robi.

Podaj punkt odniesienia i jego datę. Podaj definicje i zaznacz te, które się zmieniły. Przedstaw liczebności kohort obok wartości procentowych. Wymień równoczesne zmiany, które mogłyby wyjaśniać obserwowany ruch, i wskaż, których nie można wykluczyć. Opisz, co się nie sprawdziło, ponieważ wdrożenie, w którym wszystko się poprawiło, jest wdrożeniem, w którym niczego nie zmierzono. A tam, gdzie formułowane jest twierdzenie o stronniczości, podaj grupy, metodę, liczebności komórek i to, co pozostaje niezmierzone.

Ten dokument jest mniej imponujący niż obecna forma i znacznie bardziej użyteczny. Jest też jedyną wersją, która przetrwa lekturę osoby zainteresowanej tym, czy jest prawdziwy.

Podsumowanie

Studium przypadku jest podstawowym artefaktem dowodowym w zakupach technologii kadrowych dla przedsiębiorstw i jest najsłabszą formą dowodu stosowaną rutynowo w praktyce zawodowej. Dotyczy jednej organizacji, jest niekontrolowane, oparte na samodzielnym raportowaniu, definiowane retrospektywnie i publikowane przez stronę zainteresowaną wynikiem.

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.

Bibliografia

- Federal Trade Commission, *In the Matter of accessiBe, Inc.*, File No. 2223156, skarga i proponowane postanowienie ogłoszone 3 stycznia 2025 r.
- Fuller, Joseph, i in. *Skills-Based Hiring: The Long Road from Pronouncements to Practice*. Harvard Business School Project on Managing the Future of Work oraz The Burning Glass Institute, luty 2024.
- Tytuł VII Ustawy o prawach obywatelskich z 1964 r., z późniejszymi zmianami, 42 U.S.C. § 2000e-2, w tym § 2000e-2(l).
- Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. Part 1607, mające zastosowanie do procedur selekcji wpływających na zatrudnianie, awansowanie i inne możliwości zatrudnienia.

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.

Jak cytować

Grizzle, H. M. (2025, May 23). Cztery liczby i brak kontrfaktualności: analiza studium przypadku sztucznej inteligencji w zasobach ludzkich. Novara Consulting Group. https://www.novaracg.com/pl/2025/05/23/four-numbers-and-no-counterfactual-reading-the-artificial-intelligence-case-study-in-human-resources/