Zarządzanie Innowacyjne w Gospodarce i Biznesie nr 2(43)/2026
ISSN 2391-5129

Bartosz Rzętkiewicz https://orcid.org/0000-0002-9395-8909
Wyższa Szkoła Biznesu i Nauk o Zdrowiu w Łodzi
e-mail: bartosz.rzetkiewicz@wsbinoz.pl

Problem jakości wiedzy generowanej przez AI w decyzjach menedżerskich

The problem of the quality of AI-generated knowledge in managerial decision-making

https://doi.org/10.25312/ziwgib.909

Streszczenie

Artykuł analizuje jakość wiedzy generowanej przez sztuczną inteligencję w decyzjach menedżerskich. Punktem wyjścia jest teza, że użyteczność decyzyjna odpowiedzi tworzonych przez GenAI nie wynika z ich stylistycznej płynności, lecz z nadzoru epistemicznego, który łączy wynik działania modelu ze źródłami, kryteriami, kontekstem, aktualnością oraz odpowiedzialnością. Część empiryczna ma charakter porównawczego studium wielu przypadków i opiera się na publicznie udokumentowanych wdrożeniach: IBM AskHR w obszarze HR, BMW Group Offer Analyst w zakupach, AI @ Morgan Stanley Assistant i Debrief w doradztwie finansowym, ConsumerIQ w The Estée Lauder Companies oraz Lilli w McKinsey. Analizę uzupełnia warstwa kontrolna zaczerpnięta z badań eksperymentalnych nad GenAI w pracy wiedzochłonnej, zwłaszcza eksperymentu z konsultantami Boston Consulting Group dotyczącego tak zwanej poszarpanej granicy techa nologicznej. W artykule rozwinięto ramę JWG-AI, która obejmuje sześć wymiarów jakości wiedzy: trafność merytoryczną, możliwość prześledzenia dowodów, kompletność względem kryteriów, adekwatność kontekstową, aktualność czasową i bezpieczeństwo normatywne. Wyniki wskazują, że wysoka jakość wdrożeń GenAI zależy mniej od samej mocy modelu, a bardziej od kuratorowanych korpusów, dyscypliny wyszukiwania, jawnych kryteriów, rutyn ewaluacyjnych, ścieżek eskalacji i odpowiedniego nadzoru człowieka.

Słowa kluczowe: generatywna sztuczna inteligencja, decyzje menedżerskie, jakość wiedzy, nadzór epistemiczny

Abstract

This article investigates the quality of knowledge generated by generative artificial intelligence in managerial decision-making. It advances the argument that decision usefulness does not arise from the linguistic fluency of AI outputs, but from the epistemic governance arrangements that connect outputs with sources, criteria, context, temporal validity and accountability. The empirical part is designed as a comparative multiple-case study based on publicly documented GenAI deployments: *IBM AskHR* in human resources, *BMW Group Offer Analyst* in procurement, *AI @ Morgan Stanley Assistant* and *Debrief* in a regulated advisory and compliance context, *ConsumerIQ* at The Estée Lauder Companies in marketing intelligence, and McKinsey's *Lilli* in consulting. The analysis is strengthened by a control layer drawn from experimental studies on GenAI and knowledge work, especially the BCG consultant experiment on the jagged technological frontier. The paper develops the JWG-AI framework and operationalizes six dimensions of AI-generated knowledge quality: substantive relevance, evidential traceability, criterion completeness, contextual fit, temporal currency, and normative safety. The findings indicate that high-quality use of GenAI is less a matter of model capability alone than of curated corpora, retrieval discipline, visible decision criteria, evaluation routines, escalation paths and calibrated human review.

Keywords: generative AI, managerial decision-making, knowledge quality, epistemic governance

Rozwój generatywnej sztucznej inteligencji (GenAI) przesuwa w naukach o zarządzaniu i jakości punkt ciężkości z samej automatyzacji czynności poznawczych na problem wiarygodności przesłanek, na których menedżer opiera decyzję[1]. Nie chodzi więc wyłącznie o to, czy system tworzy odpowiedź szybko i stylistycznie poprawnie, lecz o to, czy wytworzony artefakt poznawczy może zostać obroniony jako element racjonalnego procesu decyzyjnego.

Menedżer nie podejmuje decyzji na podstawie „tekstu" jako takiego. Korzysta z propozycji interpretacji, klasyfikacji, porównania lub rekomendacji, które dopiero po weryfikacji mogą uzyskać status wiedzy organizacyjnej. To rozróżnienie jest zasadnicze, to znaczy, że płynność językowa jest własnością komunikatu, natomiast jakość wiedzy jest własnością całego układu socjotechnicznego, w którym komunikat powstaje, jest sprawdzany i zostaje powiązany z odpowiedzialnością za decyzję.

[1] W artykule pojęcie wiedzy generowanej przez AI nie oznacza autonomicznej wiedzy posiadanej przez system, lecz element poznawczy wytworzony przez GenAI, który może zostać włączony do procesu wiedzy organizacyjnej dopiero po ludzkiej i proceduralnej walidacji.

Wkład artykułu do rozważań polega na rozwinięciu ramy JWG-AI, wzbogaceniu jej o autorską skalę oceny nadzoru epistemicznego[2] oraz zestawieniu publicznie udokumentowanych wdrożeń GenAI z wynikami eksperymentów dotyczących pracy wiedzochłonnej. W ten sposób analiza wychodzi poza prostą narrację o wzroście produktywności i kieruje uwagę ku ekonomii błędu, czyli kosztowi decyzji, która została przyspieszona, ale osłabiona poznawczo.

Wprowadzenie

Generatywna sztuczna inteligencja (GenAI) weszła do organizacji przede wszystkim jako narzędzie przyspieszania pracy wiedzochłonnej. W codziennych procesach funkcjonowania firmy streszcza dokumenty, porównuje oferty, odpowiada na pytania pracowników, przygotowuje notatki po spotkaniach, syntetyzuje raporty rynkowe i wspiera tworzenie wariantów strategicznych (Csaszar, Ketkar, Kim, 2024). W praktyce menedżerskiej jej znaczenie jest jednak głębsze niż prosta automatyzacja tekstu. W naszych czasach GenAI zaczyna uczestniczyć w wytwarzaniu przesłanek decyzji (Agrawal, Gans, Goldfarb, 2018; Shrestha, Ben-Menahem, von Krogh, 2019). Jeżeli przesłanka jest błędna, niepełna, nieaktualna albo pozbawiona źródeł, szybkość jej uzyskania nie jest korzyścią organizacyjną, lecz mechanizmem przyspieszonej reprodukcji błędu.

W naukach o zarządzaniu i jakości to przesunięcie ma konsekwencje teoretyczne. Klasyczna problematyka jakości danych i informacji koncentrowała się na dokładności, kompletności, aktualności, spójności oraz użyteczności w określonym kontekście (Wang, Strong, 1996; Lee i in., 2002). GenAI komplikuje tę perspektywę. Nie zwraca danych, a wytwarza ustrukturyzowane wypowiedzi, argumenty, interpretacje i rekomendacje. Ich ryzyko polega na tym, że mogą być jednocześnie komunikacyjnie przekonujące i o słabym wymiarze epistemicznym. Innymi słowy, system może produkować formę wiedzy, która przypomina rezultat ekspercki, zanim przejdzie procedury właściwe wiedzy organizacyjnej.

Problem ten nie jest abstrakcyjny. W decyzjach HR błąd może oznaczać niezgodną z polityką lub prawem pracy odpowiedź udzieloną pracownikowi. W zakupach może prowadzić do wyboru dostawcy na podstawie niepełnego porównania ofert. W *compliance* i doradztwie finansowym może naruszyć reguły komunikacji z klientem albo wymogi dokumentacyjne. W marketingu może przekształcić chwilowy szum rynkowy w strategiczną hipotezę. W consultingu zaś może utrwalić schemat interpretacyjny dominujący w dotychczasowym korpusie wiedzy firmy. Każdy z tych przypadków dotyczy

[2] Nadzór epistemiczny oznacza zorganizowaną kontrolę wiarygodności wiedzy używanej w decyzji – jej źródeł, uzasadnienia, kompletności, aktualności i zgodności z kontekstem organizacyjnym. Termin nawiązuje do koncepcji *epistemic vigilance* Dana Sperbera i współautorów, którzy wskazują, że komunikacja wymaga mechanizmów ochrony przed błędnym poinformowaniem, a zarazem podkreślają: „vigilance […] is not the opposite of trust; it is the opposite of blind trust" (Sperber i in., 2010: 363–364). W odniesieniu do GenAI oznacza to nie nieufność wobec systemu, lecz instytucjonalne zabezpieczenie przed bezrefleksyjnym przyjęciem jego odpowiedzi jako wiedzy.

innego rodzaju szkody, acz wszystkie prowadzą do wspólnego pytania – kiedy odpowiedź GenAI zasługuje na to, by stać się przesłanką decyzji menedżerskiej?

Literatura empiryczna pokazuje, że GenAI może wyraźnie zwiększać produktywność. Shakked Noy i Whitney Zhang poprzez eksperyment z zadaniami pisarskimi wykazali, że dostęp do programu ChatGPT skracał czas pracy i podnosił jakość wyników, szczególnie u słabszych wykonawców (Noy, Zhang, 2023). Erik Brynjolfsson, Danielle Li i Lindsey Raymond udowodnili w badaniu, w którym brało udział 5000 pracowników obsługi klienta, że asystent konwersacyjny zwiększał produktywność, w tym najmocniej wspierał osoby mniej doświadczone (Brynjolfsson, Li, Raymond, 2025). Wyniki te nie oznaczają jednak, że GenAI automatycznie poprawia jakość decyzji. Eksperyment z konsultantami BCG wskazuje raczej na pewną ambiwalencję. W zadaniach mieszczących się w możliwościach modelu uzyskiwano wzrost jakości i szybkości, natomiast poza tak zwaną *jagged technological frontier*[3] użycie AI obniżało poprawność rozwiązań (Dell'Acqua i in., 2026). W tejże trudności mieści się zasadnicza luka badawcza. Dotychczasowa debata zbyt często traktuje GenAI jako technologię produktywną, podczas gdy dla nauk o zarządzaniu i jakości równie ważne wydaje się pytanie o warunki uznania wyniku pracy AI za wiedzę organizacyjną. Artykuł rozwija tę perspektywę z punktu widzenia koncepcji nadzoru epistemicznego (Alasuutari, Qadir, 2014).

Celem artykułu jest identyfikacja mechanizmów podnoszących jakość wiedzy tworzonej przez GenAI w decyzjach menedżerskich. Pytanie badawcze brzmi: jakie konfiguracje nadzoru epistemicznego sprawiają, że wyniki działania GenAI stają się poznawczo użyteczne i decyzyjnie wiarygodne w różnych domenach organizacyjnych? Odpowiedź oparto na porównawczym studium pięciu publicznie udokumentowanych wdrożeń oraz na zestawieniu ich z wynikami badań eksperymentalnych nad GenAI w pracy wiedzochłonnej.

Wkład artykułu sytuuję na styku teorii zarządzania wiedzą, badań nad jakością informacji oraz refleksji nad organizacyjnym użyciem generatywnej sztucznej inteligencji. Po pierwsze, artykuł proponuje ramę JWG-AI, która pozwala uchwycić jakość wiedzy generowanej przez AI nie jako cechę samego tekstu, lecz jako rezultat relacji między modelem, korpusem źródłowym, strukturą zadania, procedurą weryfikacji i odpowiedzialnością decydenta. Po drugie, opracowana macierz kodowania przypadków umożliwia analityczne przełożenie publicznie udokumentowanych wdrożeń GenAI na profil nadzoru epistemicznego, a więc na układ mechanizmów decydujących o tym, czy wygenerowana odpowiedź jest możliwa do sprawdzenia, kompletna, kontekstowo adekwatna i normatywnie bezpieczna. Po trzecie, artykuł rozwija propozycje teoretyczne dla przyszłych badań empirycznych, zwłaszcza eksperymentów menedżerskich i studiów terenowych, w których przedmiotem analizy byłaby nie tylko produktywność użytkownika, lecz również koszt błędu poznawczego, kalibracja zaufania oraz jakość uzasadnienia decyzji. W konsekwencji

[3] Pojęcie *jagged technological frontier* oznacza nierównomierną, trudną do przewidzenia granicę kompetencji AI: zadania podobne dla człowieka pod względem trudności mogą być przez GenAI wykonywane bardzo dobrze albo wyraźnie gorzej. Fabrizio Dell'Acqua i jego współpracownicy w badaniu z udziałem 758 konsultantów BCG wykazali, że AI poprawiała produktywność i jakość pracy w zadaniach mieszczących się w jej aktualnych możliwościach, lecz obniżała trafność rozwiązań w zadaniu znajdującym się poza tą granicą. Dlatego termin ten nie opisuje ogólnej „mocy" AI, lecz jej nieciągłą i kontekstowo zależną użyteczność w pracy wiedzochłonnej.

GenAI zostaje tu ujęta jako infrastruktura epistemiczna organizacji, czyli narzędzie współwytwarzania wiedzy, którego wartość zależy od stopnia, w jakim organizacja potrafi nie tylko projektować, ale również kontrolować i audytować warunki jego użycia. Tak ujęty temat może stać się fundamentem działań teoretycznych oraz praktycznych.

Od jakości informacji do jakości wiedzy generowanej przez AI

Pojęcie jakości w zarządzaniu ma długą tradycję, lecz w analizie GenAI wymaga jednak doprecyzowania. Jakość danych rozumiana jest zwykle wielowymiarowo. Obejmuje dokładność, kompletność, spójność, aktualność, dostępność i adekwatność do celu użycia (Wand, Wang, 1996; Wang, Strong, 1996). Yang Lee i współautorzy podkreślają, że jakość informacji jest nie tylko własnością obiektu informacyjnego, ale również relacją pomiędzy informacją a użytkownikiem, zadaniem i kontekstem. Ta relacyjność staje się szczególnie istotna w przypadku GenAI. Produkt czy rezultat modelu nie jest prostą reprezentacją bazy danych. Staje się oto wytworzoną odpowiedzią, która selekcjonuje, porządkuje i językowo scala różne elementy wiedzy (Weick, 1995; Lee i in., 2002).

Wiedza organizacyjna w literaturze zarządzania wiedzą nie jest redukowalna do informacji. Thomas Davenport i Laurence Prusak ujmowali ją jako mieszaninę doświadczenia, wartości, informacji kontekstowej i eksperckiego wglądu, która tworzy ramę oceny nowych doświadczeń (Davenport, Prusak, 1998). Ikujiro Nonaka wskazywał na dynamikę przekształceń między wiedzą *tacit* i *explicit* (Nonaka, 1994), a Maryam Alavi i Dorothy Leidner podkreślały znaczenie systemów wiedzy jako infrastruktury tworzenia, przechowywania, transferu i zastosowania wiedzy (Alavi, Leidner, 2001). GenAI ingeruje w każdy z tych etapów. Przyspiesza wyszukiwanie, upraszcza transfer, zwiększa dostępność wiedzy skodyfikowanej, ale przy okazji może osłabić refleksję nad jej pochodzeniem i zakresem obowiązywania. Efekt ten jest oczywiście wytłumaczalny także na poziomie psychologicznym. To jednak wykracza poza ramy niniejszego artykułu.

W decyzjach menedżerskich wiedza ma charakter pragmatyczny. Nie wystarcza, że komunikat jest poprawny w sensie ogólnym. Musi być trafny wobec konkretnego problemu, jurysdykcji, klienta, procedury i potencjalnej konsekwencji. Simonowska koncepcja ograniczonej racjonalności przypomina, że organizacje nie maksymalizują decyzji w warunkach pełnej informacji, lecz konstruują wystarczające rozwiązania przy ograniczonych zasobach poznawczych (Simon, 1979). GenAI może poszerzać przestrzeń poszukiwania, ale może również wprowadzać nowe ograniczenie. Może je skonstatować w następujący sposób: użytkownik otrzymuje odpowiedź, której podstaw nie potrafi szybko zweryfikować – nie potrafi i najczęściej nie chce. W nowszej literaturze problem ten bywa ujmowany jako ograniczenie epistemiczne, a nie jedynie informacyjne.

Badania nad AI w organizacjach dostarczają dwóch istotnych punktów odniesienia (Faraj, Pachidi, Sayegh, 2018; Jarrahi, 2018). Pierwszy dotyczy relacji automatyzacji i augmentacji. Sebastian Raisch i Sebastian Krakowski pokazali, że w zarządzaniu te dwa tryby nie są rozłączne. Automatyzacja pewnych etapów może zwiększać zdolności

człowieka, ale może też przesuwać odpowiedzialność i zubożać kompetencje interpretacyjne. Drugi dotyczy zaufania (Raisch, Krakowski, 2021). John Lee i Katrina See oraz Ella Glikson i Anita Woolley wskazują, że celem nie powinno być maksymalne zaufanie do automatyzacji, lecz zaufanie skalibrowane odpowiednio do kompetencji systemu i ryzyka związanego z zadaniem. GenAI, z racji płynnego języka, szczególnie łatwo prowokuje nadmierną ufność. Ujmując to kolokwialnie i lekko humorystycznie, diabeł tkwi w języku (Lee, See, 2004; Glikson, Woolley, 2020).

Na potrzeby artykułu jakość wiedzy generowanej przez AI definiuję jako stopień, w jakim wytworzony przez system GenAI artefakt poznawczy – odpowiedź, synteza, porównanie, notatka lub rekomendacja – może zostać użyty w procesie decyzyjnym po weryfikacji źródeł, kryteriów, kontekstu, aktualności i oszacowaniu normatywnego ryzyka. Definicja ta w sposób celowy nie przypisuje jakości samemu modelowi. Jakość jest tu własnością relacji, tj. modelu, korpusu, zapytania, interfejsu, procedur walidacyjnych i odpowiedzialności człowieka.

Zaproponowana rama JWG-AI obejmuje sześć wymiarów. Trafność merytoryczna oznacza, że odpowiedź rozwiązuje problem, a nie jedynie go elegancko parafrazuje. Możliwość dowodowego prześledzenia oznacza możliwość wskazania źródeł i fragmentów, z których wynika twierdzenie. Kompletność wobec kryteriów dotyczy objęcia pełnego zestawu warunków decyzyjnych. Adekwatność kontekstowa oznacza dopasowanie do organizacji, branży, jurysdykcji, klienta lub rynku. Aktualność czasowa odnosi się do kontroli wersji dokumentów i danych. Bezpieczeństwo normatywne obejmuje zgodność z prawem, politykami, etyką, poufnością i zasadami niedyskryminacji (*OECD AI Principles overview*, 2024).

Tab. 1. Rama JWG-AI. Operacjonalizacja jakości wiedzy generowanej przez AI

Wymiar jakości	**Pytanie diagnostyczne**	**Typowe ryzyko**	**Mechanizm kontroli**
Trafność merytoryczna	Czy odpowiedź rozwiązuje właściwy problem decyzyjny?	Płynna, lecz powierzchowna parafraza, pozorna eksperckość	Walidacja ekspercka, testy zadaniowe, porównanie z decyzjami referencyjnymi
Weryfikowalność źródłowa	Czy użytkownik widzi źródła, cytaty, wersje dokumentów lub dowody?	Halucynacja, brak możliwości przeprowadzania audytu, niemożność obrony decyzji	RAG, linki do dokumentów, cytowania, rejestr wykorzystanych źródeł
Kompletność kryteriów	Czy ujęto wszystkie istotne kryteria i warianty?	Pominięcie warunku granicznego lub kryterium wykluczającego	Checklisty, macierze porównawcze, jawne kryteria punktacji

Wymiar jakości	Pytanie diagnostyczne	Typowe ryzyko	Mechanizm kontroli
Adekwatność kontekstowa	Czy odpowiedź uwzględnia branżę, jednostkę, klienta i jurysdykcję?	Odpowiedź ogólna, nieprzystająca do lokalnych realiów	Kuratorowane4 korpusy źródłowe, profile użytkownika, reguły lokalne, kontekstowo osadzone szablony zapytań
Aktualność czasowa	Czy system korzysta z obowiązującej wersji polityki, raportu lub danych?	Użycie dokumentu archiwalnego, starzenie się wiedzy	Kontrola wersji, datowanie dokumentów, cykle odświeżania korpusu
Bezpieczeństwo normatywne	Czy wynik spełnia wymogi prawne, etyczne, poufnościowe i antydyskryminacyjne?	Naruszenie *compliance*, bias, wyciek informacji, nieuprawniona porada	Role dostępu, audyty, testowanie kontradyktoryjne systemu AI (ang. r*ed teaming*)[5], ścieżki eskalacji, weryfikacja ekspercka przez człowieka (ang. *human review*)

Źródło: opracowanie własne na podstawie literatury i publicznych materiałów źródłowych wskazanych w bibliografii.

Metoda badania i źródła danych

Badanie ma charakter eksploracyjnego, porównawczego studium wielu przypadków. Wybór tej strategii jest uzasadniony nowością badanego zjawiska, niedostępnością porównywalnych danych ilościowych dla większości wdrożeń GenAI oraz potrzebą uchwycenia konfiguracji organizacyjnych, a nie samego działania modelu językowego. Zgodnie z logiką studium przypadku (Eisenhardt, 1989; Eisenhardt, Graebner, 2007; Gibbert, Ruigrok, Wicki, 2008; Yin, 2018) celem nie jest uogólnienie statystyczne, lecz uogólnienie analityczne. Jest to zatem identyfikacja mechanizmów, które mogą być następnie testowane w innych kontekstach.

[4] *Źródło kuratorowane* rozumiane jest w artykule jako źródło podlegające kuracji merytorycznej. Oznacza dokument, zbiór danych, procedurę, raport, regulację, komentarz ekspercki lub inny zasób wiedzy, który może zostać oceniony pod względem autorstwa, aktualności, wiarygodności, relewantności, zgodności z politykami organizacji oraz przydatności dla określonego zadania decyzyjnego.

[5] *Testowanie kontradyktoryjne systemu AI* (ang. *red teaming*) oznacza zorganizowaną procedurę sprawdzania systemu sztucznej inteligencji poprzez celowe poszukiwanie jego błędów, podatności, zachowań niepożądanych, odpowiedzi dyskryminacyjnych, niebezpiecznych lub niezgodnych z założonym sposobem użycia. W ujęciu NIST jest to „structured testing effort to find flaws and vulnerabilities in an AI system", prowadzone często w środowisku kontrolowanym i we współpracy z twórcami systemu. W kontekście jakości wiedzy generowanej przez GenAI testowanie kontradyktoryjne nie służy więc wyłącznie bezpieczeństwu technicznemu, lecz także ocenie, czy model nie wytwarza pozornie wiarygodnych, ale błędnych, stronniczych albo normatywnie ryzykownych przesłanek decyzji menedżerskiej.

Dobór przypadków miał charakter celowy. Wybrano wdrożenia reprezentujące różne domeny decyzyjne i różne profile możliwej szkody. Są to HR (ang. *human resources*), zakupy, *compliance*/doradztwo regulowane, analityka wiedzy rynkowo-marketingowej oraz consulting. Warunkiem włączenia do próby było publiczne udokumentowanie wdrożenia przez organizację, partnera technologicznego albo wiarygodny materiał źródłowy. Ostateczny zestaw obejmuje IBM AskHR, BMW Group Offer Analyst, AI @ Morgan Stanley Assistant i Debrief, ConsumerIQ w The Estée Lauder Companies oraz Lilli w McKinsey. Dodatkowo wprowadzono przypadek kontrolny o charakterze eksperymentalnym – badanie Dell'Acqua i współautorów (2026) przeprowadzone z konsultantami BCG. Nie jest ono (jak już powiedziano) wdrożeniem organizacyjnym w takim samym sensie jak pozostałe przypadki, ale wydaje się, że pozwala sprawdzić, czy mechanizmy obserwowane w studiach wdrożeniowych odpowiadają wynikom kontrolowanego eksperymentu.

Za podstawową jednostkę analizy przyjęto rezultat poznawczy wytwarzany przez GenAI w ramach konkretnego procesu organizacyjnego. Oznacza to, że przedmiotem porównania nie były same systemy informatyczne ani ogólne strategie wdrożeniowe, lecz te formy wiedzy, które po wygenerowaniu przez model mogły zostać użyte jako przesłanki działania menedżerskiego. W analizowanych przypadkach obejmowały one odpowiedzi na pytania pracowników, porównania ofert dostawców, odpowiedzi doradcze oparte na wewnętrznych zasobach wiedzy organizacyjnej, notatki po spotkaniach, syntezy wiedzy o konsumentach i trendach rynkowych oraz syntezy wiedzy projektowej. Każdy przypadek opisano według jednolitego schematu analitycznego uwzględniającego rodzaj źródeł, typ zadania, dostępne miary efektu, zastosowane mechanizmy kontroli jakości, procedury kierowania spraw do dalszej oceny oraz zasadnicze ryzyka poznawcze, które mogły wpływać na wiarygodność decyzji podejmowanej z wykorzystaniem odpowiedzi GenAI. Kodowanie przeprowadzono na skali 0–3 dla sześciu wymiarów JWG-AI, tj. 0 – brak publicznie uchwytnego mechanizmu, 1 – mechanizm szczątkowy lub deklaratywny, 2 – mechanizm widoczny (ale częściowy), 3 – mechanizm silny, zintegrowany z procesem i możliwy do uzasadnienia na podstawie źródeł.

Takie podejście ma oczywiście ograniczenia. Materiały korporacyjne mają charakter selektywny i zwykle eksponują sukces wdrożenia. Nie dają pełnego wglądu w liczbę błędów, odrzucone odpowiedzi, koszty utrzymania, napięcia organizacyjne ani realną jakość decyzji po użyciu narzędzia. Z tego powodu w artykule nie twierdzę, że opisane wdrożenia są empirycznie najlepsze w swoich branżach. Twierdzę natomiast, że są wystarczająco dobrze udokumentowane, aby zrekonstruować ich architekturę nadzoru epistemicznego i porównać ją między domenami.

Kryterium rzetelności zapewniono przez triangulację źródeł, czyli zestawienie materiałów firmowych z literaturą naukową, standardami zarządzania ryzykiem AI (*ISO/IEC 42001:2023. Information technology – Artificial intelligence – Management system*, 2023; *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, 2023) oraz wynikami eksperymentów. Wartości liczbowe przy przypadkach przytaczano tylko wtedy, gdy były jawnie podane w źródle. Autorskie wyniki punktowania nie są miarą skuteczności ekonomicznej ani niezależnym audytem. Są bardziej jakościową oceną dojrzałości mechanizmów kontroli wiedzy ujawnionych w dostępnych materiałach.

Tab. 2. Dobór przypadków, źródła i jawne dane empiryczne wykorzystane w analizie

Przypadek	Obszar	Artefakt wiedzy	Dane/cechy wdrożenia	Podstawowe źródła dotyczące przypadku
IBM AskHR	HR	Odpowiedzi na pytania pracowników, obsługa zadań HR, personalizowane wskazania proceduralne	Ponad 80 automatyzowanych zadań HR, ponad 2,1 mln rozmów rocznie, dostęp do około 7000 stron polityk firmy, ponad 1 mln transakcji HR w 2024 roku, 55,1 tys. informacji zwrotnych, 99% adopcji wśród menedżerów, 94-procentowy wskaźnik samodzielnej obsługi zapytań przez system AI[6], 40% redukcji kosztów operacyjnych HR w cztery lata.	*Transforming HR support with agentic AI*, 2025
BMW Group Offer Analyst	Zakupy	Porównanie ofert wobec zapytania ofertowego (ang. *Request for Proposal*, RfP)[7], dokumentów i kryteriów standardowych/ *ad hoc*	Proces obejmuje wczytanie do systemu dokumentacji zapytania ofertowego oraz formularzy cenowych, ekstrakcję informacji, analizę według kryteriów standardowych i dopasowanych. BMW wskazuje też w 2025 roku wdrożenie narzędzi GenAI w zakupach, w tym Offer Analyst, Knowledge Navigator i Tender Assistant.	*Revamping procurement operations with generative AI-powered Offer Analyst solution on AWS*, 2024; *Greater efficiency and productivity with artificial intelligence – generative AI in BMW Group Purchasing*, 2025

[6] IBM deklaruje, że IBM AskHR osiągnął 94-procentowy wskaźnik samodzielnej obsługi typowych pytań (ang. *containment rate*), co oznacza, że zdecydowana większość rutynowych interakcji była zamykana przez system bez eskalacji do pracownika HR. W ujęciu jakościowym wskaźnik ten należy jednak traktować jako miarę efektywności operacyjnej, a nie jako samodzielny dowód trafności lub kompletności wiedzy generowanej przez AI.

[7] RfP oznacza zapytanie ofertowe kierowane do potencjalnych dostawców w celu uzyskania porównywalnych propozycji handlowych, technicznych i organizacyjnych. Biorąc pod uwagę systemy GenAI stosowane w zakupach, „wczytanie dokumentacji RfP oraz formularzy cenowych" oznacza przekazanie systemowi dokumentów zawierających wymagania zamawiającego, kryteria oceny oraz dane cenowe oferentów. Jakość analizy AI zależy w tym przypadku od kompletności tych dokumentów, jednoznaczności kryteriów oraz możliwości powiązania rekomendacji systemu z konkretnymi fragmentami zapytania ofertowego i formularzy cenowych.

Przypadek	Obszar	Artefakt wiedzy	Dane/cechy wdrożenia	Podstawowe źródła dotyczące przypadku
AI @ Morgan Stanley Assistant / Debrief	Doradztwo finansowe i zgodność regulacyjna (ang. *compliance*)	Odpowiedzi oparte na wewnętrznym repozytorium wiedzy organizacji (tzw. korpusie firmowym)[8], syntezy, notatki i szkice komunikacji po spotkaniach	Odpowiedzi generowane wyłącznie na podstawie wewnętrznych zasobów wiedzy Morgan Stanley Wealth Management i linkowane do dokumentów, 98% adopcji Assistant w zespołach doradców, wzrost dostępu do dokumentów z 20% do 80%, Debrief za zgodą klienta tworzy notatki, identyfikuje ustalenia do realizacji wynikające ze spotkania i szkice e-maili do dalszej redakcji przez doradcę, system ewaluacji jakości działania GenAI, zestaw testów regresyjnych, zerowa retencja danych[9], rozumiana jako tryb nieutrwalania treści klienta w logach dostawcy, z zastrzeżeniem ograniczeń technicznych i kontraktowych.	*Morgan Stanley Wealth Management Announces Key Milestone in Innovation Journey with OpenAI*, 2023; *Morgan Stanley Research Announces AskResearchGPT Powered by OpenAI*, 2024; *Morgan Stanley uses AI evals to shape the future of financial services*, 2025

[8] Odpowiedzi generowane na podstawie korpusu wiedzy organizacyjnej oznaczają rezultaty działania systemu GenAI, które są tworzone nie z ogólnej wiedzy modelu ani z otwartego internetu, ale z uprzednio wyselekcjonowanego, zweryfikowanego i kontrolowanego zbioru dokumentów należących do organizacji. Taki korpus może obejmować procedury, polityki wewnętrzne, raporty, analizy, instrukcje, komentarze eksperckie, regulaminy, dokumentację klienta lub inne zasoby skodyfikowanej wiedzy. W ujęciu jakościowym istotne jest to, że odpowiedź powinna dać się powiązać z konkretnymi fragmentami tego korpusu. Wzmacnia to możliwość jej weryfikacji źródłowej, ogranicza ryzyko halucynacji i umożliwia audyt decyzji podjętej na jej podstawie.

[9] Zerowa retencja danych (ang. *zero data retention*) oznacza tryb przetwarzania, w którym treści klienta, takie jak zapytania i odpowiedzi systemu, nie są utrwalane w logach monitorowania nadużyć po stronie dostawcy usługi. Nie należy jednak utożsamiać tego pojęcia z absolutnym brakiem przetwarzania lub z identycznym statusem wszystkich funkcji systemu. W praktyce zakres zerowej retencji zależy od polityki dostawcy, rodzaju usługi, tak zwanego endpointu, umowy oraz wymogów bezpieczeństwa i zgodności regulacyjnej. W zastosowaniach menedżerskich, zwłaszcza w sektorach regulowanych, zerowa retencja danych stanowi mechanizm ograniczania ryzyka ujawnienia informacji poufnych, lecz nie zastępuje wewnętrznych procedur klasyfikacji danych, kontroli dostępu i audytu użycia AI.

Przypadek	Obszar	Artefakt wiedzy	Dane/cechy wdrożenia	Podstawowe źródła dotyczące przypadku
ConsumerIQ – The Estée Lauder Companies	Analityka wiedzy o rynku i konsumentach[10] (ang. *marketing intelligence*)	Synteza wiedzy o konsumentach[11], trendów i dokumentów w różnych formatach	Niemal 25 marek w około 150 krajach, dokumenty w formacie pdf, PowerPoint i arkuszach, centralizacja insightów, skrócenie czasu dostępu do danych z tygodni do minut, odpowiedzi z cytowaniami do walidacji źródeł.	*The Estée Lauder Companies and Microsoft Copilot Studio transform intelligence*, 2025; *ELC Partners With Microsoft to Create ConsumerIQ Tool for Data-Driven Insights*, 2025
McKinsey Lilli	Consulting	Wyszukiwanie, synteza wiedzy firmowej, wskazywanie ekspertów i materiałów	Ponad 40 źródeł podlegających kuracji merytorycznej, ponad 100 tys. dokumentów i transkryptów wywiadów, sieć ekspertów w 70 krajach, odpowiedzi z linkami, 5–7 najbardziej relewantnych materiałów, data stewards, wskazania oszczędności czasu. przygotowania do spotkań do 20% w wypowiedzi użytkownika	*Meet Lilli, our generative AI tool that's a researcher, a time saver, and an inspiration*, 2023
Eksperyment BCG	Consulting/praca wiedzochłonna	Rozwiązania realistycznych zadań konsultingowych z użyciem GPT-4	Eksperyment terenowy (758 konsultantów BCG), porównanie trzech warunków pracy: bez wsparcia AI, z wykorzystaniem GPT-4 oraz z wykorzystaniem GPT-4 po instruktażu dotyczącym formułowania poleceń. W zadaniach znajdujących się w aktualnym zakresie możliwości AI użycie GPT-4 prowadziło do wykonania średnio o 12,2% większej liczby zadań, skrócenia czasu pracy o 25,1% oraz wzrostu jakości rezultatów o ponad 40%. Efekt ten załamywał się jednak poza granicą kompetencji systemu – w takim zadaniu poprawność rozwiązań była o 19 p.p. niższa niż w grupie pracującej bez AI.	Dell'Acqua i in., 2026

Źródło: opracowanie własne na podstawie literatury i publicznych materiałów źródłowych wskazanych w bibliografii.

[10] Analityka wiedzy o rynku i konsumentach oznacza zorganizowany proces pozyskiwania, porządkowania, łączenia i interpretowania informacji dotyczących klientów, konkurencji, trendów, kanałów sprzedaży, zachowań konsumenckich oraz szerszego otoczenia rynkowego. Jej celem nie jest samo gromadzenie danych, lecz przekształcanie rozproszonych sygnałów rynkowych we wnioski użyteczne dla decyzji marketingowych, produktowych i strategicznych.

[11] Synteza wiedzy o konsumentach oznacza proces integrowania i interpretowania rozproszonych danych, obserwacji oraz wniosków dotyczących potrzeb, motywacji, preferencji i zachowań konsumentów. W odróżnieniu od prostej agregacji danych synteza ta polega na nadaniu informacjom znaczenia decyzyjnego, tj. na przekształceniu wielu cząstkowych sygnałów rynkowych w spójną diagnozę użyteczną dla decyzji marketingowych, produktowych lub strategicznych. W odniesieniu do GenAI pojęcie to wymaga ostrożności. System może sprawnie łączyć rozproszone materiały, ale zarazem wzmacniać nadinterpretacje, uproszczenia lub słabo udokumentowane hipotezy dotyczące konsumentów.

Wyniki studiów przypadków

IBM AskHR. Aktualność polityk, kontekst lokalny i obsługa przypadków nietypowych jako kryteria jakości wiedzy HR

IBM AskHR jest dobrym przykładem tego, że jakość wiedzy generowanej przez AI w HR zależy od wcześniejszej dyscypliny procesowej. W opisie IBM szczególnie istotna jest zasada *transform, not transfer* – czyli, innymi słowy, wadliwych procesów nie należy po prostu przenosić do automatyzacji, lecz najpierw je przeprojektować. To zdanie ma duże znaczenie teoretyczne. Wskazuje, że jakość odpowiedzi GenAI jest pochodną jakości procesu, z którego odpowiedź wyrasta. Model nie naprawia chaosu polityk HR – co najwyżej generuje bardziej płynne odpowiedzi na podstawie chaotycznego korpusu.

W publicznie dostępnym opisie wdrożenia IBM AskHR przedstawiono szereg wskaźników obrazujących skalę i organizacyjne osadzenie systemu. Narzędzie automatyzuje ponad 80 typów zadań z obszaru HR, obsługuje rocznie ponad 2,1 mln interakcji z pracownikami, korzysta z zasobu obejmującego około 7000 stron polityk organizacyjnych, a w 2024 roku przetworzyło ponad 1 mln transakcji HR. IBM wskazuje ponadto na 55,1 tys. informacji zwrotnych zebranych w 2024 roku, 99-procentowy poziom wykorzystania narzędzia wśród menedżerów, 94-procentowy wskaźnik samodzielnej obsługi typowych pytań bez przekazania sprawy człowiekowi oraz 40-procentową redukcję kosztów operacyjnych HR w okresie czterech lat (*Transforming HR support with agentic AI*, 2025). Dane te nie dowodzą jeszcze jakości każdej pojedynczej odpowiedzi generowanej przez system. Pozwalają uchwycić trzy istotne warunki zarządzania jakością wiedzy, tj. skalę (powtarzalnego) użycia, istnienie (mierzalnego) obiegu informacji zwrotnej oraz operacyjne powiązanie systemu z (aktualizowanym) zasobem polityk organizacyjnych.

Z perspektywy ramy JWG-AI przypadek IBM AskHR pokazuje, że trafność merytoryczna odpowiedzi zależy przede wszystkim od poprawnego przypisania zapytania pracownika do właściwej domeny HR (świadczeń, wynagrodzeń, rozwoju kariery, podróży służbowych, dokumentacji pracowniczej bądź czynności menedżerskich). Szczególnego znaczenia nabiera tu aktualność wiedzy, ponieważ polityki HR podlegają oczywistym zmianom w czasie i różnicują się terytorialnie. Odpowiedź zgodna z procedurą obowiązującą w jednej jurysdykcji może być nietrafna w innej, a komunikat poprawny w poprzednim okresie rozliczeniowym może utracić ważność po zmianie regulaminu, procesu albo lokalnych przepisów prawa pracy. Równie istotny jest wymiar bezpieczeństwa normatywnego. Obejmuje on ochronę danych pracownika, niedyskryminacyjny charakter odpowiedzi oraz zgodność z regulacjami właściwymi dla danego kraju i jednostki organizacyjnej.

Z analizowanego przypadku wynika zatem, że system GenAI stosowany w HR nie powinien być projektowany jako ogólny chatbot konwersacyjny, ale jako ograniczony domeną mechanizm proceduralnego wsparcia, który działa na podstawie wcześniej zweryfikowanych i aktualnych polityk organizacyjnych. O jego jakości decyduje nie tylko zdolność do udzielania odpowiedzi na rutynowe pytania, ale także zdolność do rozpoznania granic własnej użyteczności. Dlatego model dwupoziomowy (automatyczna obsługa

spraw standardowych oraz przekazywanie przypadków złożonych, nietypowych lub wrażliwych do oceny człowieka) należy traktować nie jako rozwiązanie pomocnicze, ale jako podstawowy warunek wiarygodności systemu. Bez takiego mechanizmu sprawy wymagające uwzględnienia lokalnego prawa pracy, indywidualnej sytuacji pracownika albo oceny etyczno-organizacyjnej mogłyby zostać błędnie zakwalifikowane jako rutynowe zapytania administracyjne.

Właśnie w tym miejscu, inaczej to ujmując, ujawnia się zasadnicza różnica między automatyzacją obsługi HR a zarządzaniem jakością wiedzy HR: pierwsza skraca czas odpowiedzi, druga wymaga rozpoznania, kiedy odpowiedź nie powinna zostać udzielona automatycznie.

BMW Group Offer Analyst. Kryteria, źródła i uzasadnienie w ocenie ofert

Przypadek BMW Group Offer Analyst ukazuje odmienny od HR sposób konstytuowania jakości wiedzy generowanej przez GenAI. W procesach zakupowych zasadniczym problemem nie jest szybkie udzielenie odpowiedzi użytkownikowi. Ważne jest tu możliwie jak najbardziej kompletne, porównywalne i źródłowo weryfikowalne przeanalizowanie dokumentacji przetargowej (zapytania ofertowego, ofert dostawców, formularzy cenowych, warunków handlowych oraz wymagań technicznych i organizacyjnych). Opis wdrożenia przedstawiony przez AWS wskazuje, że proces obejmuje wczytanie do systemu dokumentacji zapytania ofertowego, następnie ofert dostawców, wydobycie informacji o oferentach, analizę według kryteriów standardowych oraz analizę według kryteriów definiowanych przez użytkownika dla konkretnego postępowania. W tym sensie podstawową cechą rozwiązania jest przekształcenie zadania konwersacyjnego w zadanie porównawczo-kryterialne: system nie tyle „odpowiada" na pytanie, ile porządkuje materiał decyzyjny według (jawnie) określonych przesłanek oceny.

Wysoka jakość analizy zakupowej powstaje tu nie przez rozszerzanie swobody modelu, lecz przeciwnie – przez jej metodyczne ograniczanie. Im silniej odpowiedź zostaje związana z dokumentami źródłowymi, kryteriami oceny i strukturą postępowania zakupowego, tym mniejsze jest ryzyko, że wynik działania GenAI przyjmie postać nieuzasadnionej sugestii lub językowo przekonującej opinii. W takiej konfiguracji rezultat pracy systemu może być traktowany jako zestawienie podlegające audytowi. Użytkownik ma możliwość sprawdzenia, które wymagania zostały spełnione, które pozostają niejednoznaczne, a które wymagają dodatkowej interpretacji. Rozwiązanie to wzmacnia w pierwszej kolejności dwa wymiary jakości wiedzy. Są to kompletność kryteriów oraz weryfikowalność źródłową uzasadnienia. Jednocześnie materiały BMW z 2025 roku sytuują Offer Analyst w szerszym programie cyfryzacji zakupów (obok takich narzędzi, jak Knowledge Navigator i Tender Assistant). Sugeruje to, że jakość analizy nie jest projektowana wyłącznie na poziomie pojedynczej aplikacji. Dokonywane jest to w ramach szerszej architektury platformowej, która obejmuje także przygotowanie użytkowników do pracy z narzędziami GenAI.

Z perspektywy ramy JWG-AI znaczenie mają dwa rozwiązania organizacyjno-poznawcze. Pierwsze polega na oddzieleniu kryteriów stałych, właściwych dla powtarzalnych

procesów zakupowych, od kryteriów specyficznych dla danego projektu, kategorii zakupowej lub postępowania. Drugie polega na aktywnym włączeniu użytkownika w definiowanie wymagań niestandardowych. Dzięki temu system nie narzuca jednej, uniwersalnej logiki oceny ofert. Umożliwia menedżerowi zakupów doprecyzowanie tych warunków, które w danym przypadku mają rzeczywiste znaczenie decyzyjne. GenAI nie zastępuje więc profesjonalnego osądu zakupowego. Porządkuje obszar do porównania, zmniejsza ryzyko pominięcia istotnych informacji i ułatwia rekonstrukcję podstaw dokonanego wyboru.

Granice tego przypadku są jednak równie istotne, jak jego potencjał. Decyzja o wyborze dostawcy nie wyczerpuje się w analizie dokumentów. Obejmuje także ocenę ryzyka relacyjnego, odporności łańcucha dostaw, historii współpracy, zdolności innowacyjnej partnera, uwarunkowań geopolitycznych oraz wiedzy ukrytej, która często nie zostaje w pełni odzwierciedlona w dokumentacji przetargowej. Offer Analyst może zatem znacząco podnieść jakość formalnej analizy ofert, ale nie powinien być interpretowany jako samodzielny mechanizm do strategicznej rekomendacji zakupowej. Jego największa wartość ujawnia się wtedy, gdy ogranicza przeoczenia, redukuje asymetrię informacyjną i zwiększa przejrzystość porównania, nie zaś wtedy, gdy miałby pozorować pełną ocenę partnerstwa biznesowego. W zakupach GenAI powinna być więc rozumiana jako narzędzie wzmacniające dyscyplinę analityczną procesu. Nie może być traktowana jako substytut odpowiedzialnego osądu menedżerskiego.

Morgan Stanley. Weryfikowalność źródłowa, ewaluacja jakości odpowiedzi i zgodność regulacyjna w doradztwie finansowym

Przypadek Morgan Stanley należy do najlepiej udokumentowanych przykładów organizacyjnego nadzoru nad wiedzą generowaną przez GenAI w wysoce regulowanym środowisku działania. Już w komunikacie z 2023 roku podkreślano, że narzędzie przeznaczone dla doradców finansowych ma wytwarzać odpowiedzi wyłącznie na podstawie wewnętrznych zasobów Morgan Stanley Wealth Management oraz odsyłać użytkownika do dokumentów źródłowych, na których oparto daną odpowiedź (*Morgan Stanley Wealth Management Announces Key Milestone in Innovation Journey with OpenAI*, 2023). W kolejnym etapie wdrożono AI @ Morgan Stanley Debrief, system sporządzający (za zgodą klienta) notatki ze spotkań, identyfikujący ustalenia do realizacji oraz przygotowujący roboczy projekt wiadomości po rozmowie. Istotne jest jednak to, że wygenerowane treści nie są traktowane jako gotowa komunikacja z klientem. Doradca zachowuje obowiązek ich sprawdzenia, redakcji i ostatecznego zatwierdzenia przed użyciem (*Morgan Stanley Research Announces AskResearchGPT Powered by OpenAI*, 2024; *Morgan Stanley Wealth Management Announces Latest Game-Changing Addition to Suite of GenAI Tools*, 2024). Rola człowieka nie ma więc charakteru pozornego ani wyłącznie formalnego. Jest ostatnim ogniwem odpowiedzialności za treść przekazywaną klientowi i za zgodność tej treści z wymogami profesjonalnego doradztwa finansowego.

Opis wdrożenia przedstawiony przez OpenAI wskazuje, że jakość systemu była budowana nie tylko przez dobór źródeł. Zaangażowany był też do tego rozbudowany system

ewaluacji odpowiedzi. Morgan Stanley testował przypadki użycia przed wdrożeniem, prowadził ewaluacje streszczeń, angażował ekspertów dziedzinowych i specjalistów odpowiedzialnych za projektowanie zapytań do oceny trafności, spójności i przydatności odpowiedzi, rozwijał ewaluacje tłumaczeń oraz doskonalił metody wyszukiwania i przywoływania właściwych źródeł. Wysokie znaczenie ma informacja o codziennym wykorzystywaniu zestawu testów regresyjnych oraz o ścisłym powiązaniu zapewniania jakości z wymogami zgodności regulacyjnej (*Morgan Stanley uses AI evals to shape the future of financial services*, 2025). Właśnie ta kombinacja – kontrolowany korpus wiedzy, systematyczna ewaluacja, testy regresyjne i nadzór zgodności – odróżnia wdrożenie w sektorze regulowanym od zwykłego użycia chatbota, który jest publicznie dostępny.

Ogólnie dostępne dane wskazują również na wysoką skalę wykorzystania rozwiązania. Morgan Stanley deklaruje, że ponad 98% zespołów doradców korzysta z narzędzi, takich jak Assistant, a dostępność dokumentów wzrosła z 20% do 80%. Komunikat dotyczący Debrief przypomina zarazem, że Assistant został w pełni udostępniony we wrześniu 2023 roku, natomiast Debrief rozszerzył funkcjonalność systemu o notowanie spotkań, streszczanie rozmów i przygotowywanie wstępnych projektów komunikacji z klientem (*Morgan Stanley Research Announces AskResearchGPT Powered by OpenAI*, 2024; *Morgan Stanley Wealth Management Announces Latest Game-Changing Addition to Suite of GenAI Tools*, 2024). Ważne i ciekawe jest przy tym to, że sama organizacja zastrzega ograniczenia wyników generowanych przez AI, w tym możliwość nieścisłości, niekompletności oraz stronniczości. Użytkownik pozostaje zobowiązany do weryfikacji treści przed ich użyciem. Z perspektywy jakości wiedzy jest to element wartościowy, a można powiedzieć, że i zasadniczy. Organizacja nie redukuje wdrożenia do narracji o produktywności. Stara się formalnie utrzymywać wymóg sprawdzenia i odpowiedzialnego użycia wyniku działania systemu.

Wniosek z tego przypadku odnosi się do związku pomiędzy weryfikowalnością źródłową a zaufaniem do wiedzy generowanej przez GenAI. Wydaje się, że w doradztwie finansowym nie wystarcza, aby odpowiedź była językowo przekonująca lub prawdopodobna. Musi być możliwa do uzasadnienia wobec klienta, regulatora oraz wewnętrznych funkcji nadzoru i zgodności. Jakość wiedzy nie sprowadza się więc do semantycznej trafności odpowiedzi. Musi obejmować możliwość odtworzenia całej sekwencji decyzyjnej – od pytania użytkownika, przez źródło i wygenerowaną odpowiedź, po ocenę ekspercką, użycie komunikatu oraz odpowiedzialność za jego konsekwencje. Przypadek Morgan Stanley pokazuje, że GenAI w finansach (i zarządzaniu nimi) wymaga technik generowania wspomaganego wyszukiwaniem oraz systemu ewaluacji jakości odpowiedzi, kontroli prywatności danych, testów regresyjnych i jednoznacznie określonej roli doradcy w procesie decyzyjnym.

Ograniczenie tego modelu wynika z tej samej cechy, która stanowi o jego bezpieczeństwie. Oparcie systemu na wewnętrznym, skodyfikowanym korpusie wiedzy może ograniczać ryzyko nieautoryzowanych rekomendacji oraz wzmacniać konserwatyzm poznawczy. System bardzo dobrze odtwarza wiedzę uznaną, zatwierdzoną i utrwaloną w organizacji. Słabiej wychwytuje sygnały peryferyjne, nowe sprzeczności, niestandardowe interpretacje lub zjawiska dopiero wyłaniające się poza głównym nurtem firmowej

ekspertyzy. W sektorze regulowanym taka ostrożność jest częściowo pożądana, ponieważ zmniejsza ryzyko komunikacji niezgodnej z politykami firmy lub wymogami nadzorczymi. Z perspektywy nauk o zarządzaniu wykazuje jednak trudność ujawniającą się między bezpieczeństwem wiedzy a jej odkrywczością. Morgan Stanley pokazuje, jak można zwiększać wiarygodność GenAI w środowisku wysokiego ryzyka. Pokazuje również, jaką cenę poznawczą może mieć daleko posunięta kontrola źródeł i procedur.

ConsumerIQ w The Estée Lauder Companies. Synteza wiedzy o konsumentach w warunkach zmienności rynku

ConsumerIQ w The Estée Lauder Companies reprezentuje ten typ zastosowania GenAI, w którym podstawową funkcją systemu nie jest obsługa powtarzalnej procedury, a integracja interpretacyjna rozproszonych zasobów wiedzy o rynku, konsumentach i trendach. ELC działa za pośrednictwem niemal 25 marek w około 150 krajach. Organizacja operuje na znacznej liczbie danych dotyczących zachowań konsumenckich, kategorii produktowych, lokalnych uwarunkowań rynkowych oraz zmian w języku i praktykach wynikających z codziennej konsumpcji. Materiały Microsoftu i ELC wskazują, że przed wdrożeniem ConsumerIQ wiedza ta była rozproszona w wielu formatach i na wielu nośnikach, obejmujących między innymi pliki pdf, prezentacje, arkusze kalkulacyjne i dokumenty analityczne. ConsumerIQ, zbudowany z wykorzystaniem Microsoft Copilot Studio oraz Azure OpenAI Service, ma porządkować te zasoby, analizować dokumenty, identyfikować wzorce i formułować rekomendacje użyteczne dla decyzji marketingowych i strategicznych (*The Estée Lauder Companies and Microsoft Copilot Studio transform intelligence*, 2025; *ELC Partners With Microsoft to Create ConsumerIQ Tool for Data-Driven Insights*, 2025).

Skrócenie czasu dostępu do istotnych danych z tygodni do minut to deklarowany efekt wdrożenia. W obszarze analityki wiedzy o rynku i konsumentach zmiana ta ma znaczenie nie tylko operacyjne, ale i poznawcze. Wartość rozpoznania rynkowego często zależy od szybkości uchwycenia zmiany – zanim stanie się ona oczywista dla konkurentów. Z perspektywy ramy JWG-AI szczególnie istotne jest jednak nie tylko samo przyspieszenie dostępu do informacji, liczy się to, w jaki sposób system wiąże wygenerowane odpowiedzi ze źródłami. Cytowania umożliwiające przejście do dokumentów źródłowych wzmacniają poziom możliwości weryfikowania formułowanych wniosków. Nie jest to element drugorzędny ani wyłącznie redakcyjny, tym bardziej w analizie trendów i zachowań konsumenckich. Źródłowe uzasadnienie odpowiedzi ogranicza ryzyko, że system przekształci krótkotrwałą modę, anegdotyczny sygnał, błąd danych wejściowych albo dominującą narrację marketingową w – pozornie tylko – prawdziwą przesłankę strategiczną.

Tutaj pojawia się jednak odmienny problem jakości niż w HR, zakupach czy doradztwie regulowanym. Odpowiedź systemu rzadko wiąże się z jednoznaczną procedurą lub pojedynczym wymogiem formalnym. Wydaje się, że częściej ma charakter hipotezy interpretacyjnej. Wskazuje, które z zachowań konsumentów mogą się utrwalać, jak zmienia się język danej kategorii, które mikrotrendy mogą oddziaływać na produkt, komunikację marki, opakowanie lub pozycjonowanie oferty. Jakość wiedzy generowanej przez GenAI

obejmuje zatem nie tylko aktualność informacji. To także zdolność do łączenia słabych sygnałów w uzasadnioną i możliwą do dalszej weryfikacji diagnozę. Im większy poziom syntezy, tym większe zarazem ryzyko wygładzenia różnic między rynkami, grupami konsumentów, markami i kontekstami kulturowymi. Ostatecznie to także większa efektywność biznesowa dla całej firmy.

Przypadek ConsumerIQ pokazuje, że w zastosowaniach marketingowych znaczenie ma nie tylko dostęp do źródeł. Widać to w obszarze różnorodności i jakości samej wiedzy. Jeżeli materiał źródłowy jest zbyt jednorodny, system może znacząco przyspieszyć analizę – nawet nie poszerzając jej zakresu poznawczego. Wysoka jakość wymaga więc kontroli tego, czy GenAI korzysta z danych globalnych oraz lokalnych, ilościowych i jakościowych, bieżących i historycznych. Równie ważne jest to, czy użytkownik potrafi odróżnić wniosek oparty na stabilnym materiale empirycznym od atrakcyjnie sformułowanej, aczkolwiek niezbyt przekonująco udokumentowanej hipotezy. W takim rozumieniu wdrożenie GenAI w marketingu wymaga szczególnie rozwiniętej kultury interpretacji. Inne, ale ważne cechy to umiejętność pracy ze źródłem, rozpoznawania niepewności, ważenia sygnałów oraz krytycznego traktowania syntetycznych odpowiedzi generowanych przez system.

McKinsey Lilli. Adekwatność kontekstowa i jakość syntezy wiedzy eksperckiej w doradztwie strategicznym

McKinsey Lilli jest przypadkiem, w którym GenAI zostaje umieszczona w samym centrum modelu profesjonalnej usługi. McKinsey pisze wprost: „Knowledge is the lifeforce of McKinsey". Jak się wydaje, to nie jest jedynie hasło wizerunkowe. W firmie doradczej jakość pracy zależy od zdolności dotarcia do właściwych analogii, ekspertów, danych, przypadków, ram interpretacyjnych. Lilli ma agregować ponad 40 kuratorowanych źródeł wiedzy, ponad 100 tys. dokumentów i transkryptów wywiadów oraz sieć ekspertów w 70 krajach.

Zgodnie z opisem McKinsey Lilli nie ogranicza się do udzielenia pojedynczej odpowiedzi na pytanie użytkownika. Po wpisaniu zapytania system identyfikuje zwykle od pięciu do siedmiu najbardziej odpowiadających materiałów, syntetyzuje ich zasadniczą treść, wskazuje odsyłacze do źródeł oraz pomaga odnaleźć ekspertów, których doświadczenie może być użyteczne w danym problemie projektowym. Jakość nie jest tu zatem rozumiana jako sprawność wygenerowania jednego tekstu. To bardziej zdolność skrócenia drogi od problemu klienta do właściwych zasobów wiedzy, dokumentów i osób. W tym oto sensie Lilli pełni funkcję organizacyjnego mechanizmu dostępu do kapitału intelektualnego firmy. Porządkuje rozproszone zasoby, ułatwia ich interpretację i przyspiesza przejście od zapytania projektowego do roboczej hipotezy rozwiązania.

Z perspektywy nadzoru nad jakością wiedzy (a z badawczej jako element poznawczy) szczególne znaczenie ma to, że wdrożenie Lilli zostało powiązane z instytucjonalnymi rolami odpowiedzialnymi za porządkowanie, ocenę i utrzymywanie zasobów źródłowych. McKinsey wskazuje, że setki osób uczestniczą w kuracji treści. Co to oznacza? Dbają o jakość danych, usuwają lub maskują informacje poufne dotyczące klientów,

opisują materiały w sposób umożliwiający ich wyszukiwanie oraz zapewniają ich przydatność dla pracy konsultantów. Ten element jest wysokiej wagi, choć często pomijany w uproszczonych dyskusjach o GenAI. Model językowy nie zastępuje bowiem organizacyjnej pracy nad wiedzą. Przeciwnie, czyni ją bardziej widoczną i strategicznie istotną. Im większa występuje zależność organizacji od systemów GenAI, tym większe jest znaczenie kuratorów zasobów wiedzy, właścicieli danych, dziedzinowych ekspertów oraz funkcji odpowiedzialnych za bezpieczeństwo, poufność i ryzyko.

Przypadek Lilli dobrze ujawnia trudności właściwe profesjonalnym organizacjom, których przewaga konkurencyjna opiera się na skodyfikowanym doświadczeniu. System bardzo skuteczny w przeszukiwaniu wcześniejszych projektów może wzmacniać analogie do rozwiązań już opisanych, uznanych i utrwalonych w korpusie wiedzy firmy. Jest to jednocześnie jego siła i ograniczenie. Siła polega na szybkim uruchomieniu doświadczenia zgromadzonego w wielu zespołach, sektorach i regionach. Ograniczenie polega na możliwości reprodukowania dominujących wzorców interpretacyjnych – a więc faworyzowania tego, co zostało już wcześniej nazwane, sklasyfikowane i uznane za wartościowe.

Dla doradztwa strategicznego ma to znaczenie zasadnicze. Najtrudniejsze problemy klientów nie zawsze wymagają odnalezienia podobnego przypadku z przeszłości. Często wymagają one rozpoznania, czy dostępna analogia jest niepełna, myląca albo pozornie podobna. Jakość wiedzy generowanej przez GenAI nie może więc polegać wyłącznie na sprawnej syntezie istniejącego zasobu. Musi obejmować również zdolność ujawniania niepewności, wskazywania luk w materiale źródłowym, sygnalizowania granic porównania, ale także – co jednak bardzo istotne – odróżniania dobrze udokumentowanej analogii od interpretacyjnej sugestii. W przypadku Lilli najważniejszy wniosek brzmiał będzie zatem następująco: GenAI może wzmacniać jakość pracy doradczej (konsultingowej) tylko wtedy, gdy dostęp do skodyfikowanej wiedzy w danej organizacji jest równoważony krytycznym namysłem nad tym, czego ta wiedza nie obejmuje.

Eksperymenty z pracą wiedzochłonną i granice produktywności

Porównawcze studia przypadków pokazują różne (i zarazem badawczo ciekawe) architektury wdrożeń, ale nie pozwalają same w sobie ocenić wpływu GenAI na jakość decyzji. Z tego powodu warto zestawić je z wynikami eksperymentów. Najważniejszy dla niniejszego artykułu jest eksperyment Dell'Acqua i współautorów przeprowadzony z 758 konsultantami BCG. Badani zostali losowo przypisani do warunków – bez dostępu do AI, z dostępem do GPT-4 albo z dostępem do GPT-4 i krótkim instruktażem promptowania (czyli przygotowania odpowiedniej instrukcji komunikacji z chatbotem). W zadaniach mieszczących się w możliwościach modelu konsultanci z AI wykonali średnio 12,2% więcej zadań, pracowali 25,1% szybciej, a jakość ich wyników była ponad 40% wyższa niż w grupie kontrolnej. Jednak w zadaniu znajdującym się poza granicą możliwości AI użytkownicy z AI byli o 19 punktów procentowych mniej skłonni do poprawnej odpowiedzi.

Wynik ma istotny wpływ na teorię jakości wiedzy. Oznacza, że GenAI może równocześnie podnosić jakość i ją niszczyć – zależnie od typu zadania. Problem nie polega więc na ogólnym pytaniu „czy AI pomaga", ale na pytaniu, czy organizacja potrafi rozpoznać granicę kompetencji systemu. Bez tej zdolności menedżer może przenieść zaufanie

wypracowane w zadaniach łatwiejszych na zadania, w których model generuje pozornie przekonujące, aczkolwiek niepoprawne rozwiązania. To właśnie stanowi sam środek rozważań o ekonomii błędu.

Podobny wniosek wynika z badań Noya i Zhang (2023), Brynjolfssona, Li i Raymond (2025) oraz Nicholasa Otisa i współautorów (2024). GenAI potrafi poprawiać wydajność i wyrównywać różnice kompetencyjne, ale efekty są heterogeniczne. W przedsiębiorczości, gdzie zadania są w dużej mierze bardziej kontekstowe i splecione z lokalnymi warunkami, sam dostęp do porad AI nie musi prowadzić do poprawy wyników ekonomicznych. Natomiast z punktu widzenia zarządzania oznacza to, że jakość wiedzy generowanej przez AI nie może być mierzona tylko czasem wykonania ani subiektywną satysfakcją użytkownika. Konieczne są miary trafności, kompletności, kalibracji pewności i skutku decyzji. To oczywiste.

Dlatego w dalszej części artykułu przypadki wdrożeniowe interpretuję przez pryzmat eksperymentalnej lekcji o granicy technologicznej. Wysoka jakość GenAI w organizacji wymaga nie tylko dostępu do modelu. Wymaga praktyk, które dają możliwość pomocy użytkownikowi w zorientowaniu się, kiedy odpowiedź jest w granicy kompetencji systemu, kiedy wymaga dodatkowego źródła, a kiedy zaś powinna zostać przekazana ekspertowi do dalszej pracy.

Porównawcza analiza nadzoru epistemicznego

Porównanie analizowanych przypadków wskazuje, że jakość wiedzy generowanej przez GenAI nie ma charakteru uniwersalnego. Przyjmuje postać domenowo zróżnicowanych konfiguracji i form kontroli. W obszarze zarządzania zasobami ludzkimi zasadnicze znaczenie mają aktualność polityk organizacyjnych, dostosowanie odpowiedzi do lokalnego kontekstu prawno-organizacyjnego oraz sprawne przekazywanie przypadków nietypowych do oceny eksperckiej. W zakupach ciężar jakości przesuwa się ku kompletności kryteriów oceny i weryfikowalności źródłowej podczas porównania ofert. W doradztwie finansowym, funkcjonującym pod presją regulacyjną, forpocztą stają się możliwość odtworzenia podstaw odpowiedzi, systematyczna ewaluacja jakości działania danego narzędzia, zgodność regulacyjna oraz odpowiedzialna weryfikacja treści przez doradcę. W marketingu jakość zależy od tempa syntezy wiedzy o rynku i konsumentach, ale też od zdolności oceny wiarygodności, różnorodności oraz aktualności źródeł, z których wyprowadzane są wnioski o trendach. W doradztwie strategicznym natomiast centralne znaczenie zyskuje organizacyjne porządkowanie wiedzy eksperckiej oraz zdolność systemu do wskazywania nie tylko gotowych odpowiedzi, ale również właściwych materiałów, źródeł i osób kompetentnych w danym problemie.

Jak już wcześniej ustalono, dla potrzeb syntezy porównawczej zastosowano skalę od 0 do 3 dla każdego z sześciu wymiarów ramy JWG-AI. Należy podkreślić, że uzyskane wyniki nie stanowią pomiaru jakości samych modeli językowych ani audytu badanych organizacji. Są one analityczną interpretacją jawnie opisanych mechanizmów zarządzania jakością wiedzy w poszczególnych wdrożeniach. Wysoka ocena oznacza, że w dostępnych

materiałach źródłowych można zidentyfikować wyraźny, powtarzalny i powiązany z procesem użycia mechanizm kontroli danego wymiaru jakości. Ocena pośrednia oznacza obecność takiego mechanizmu, lecz w formie częściowej, ogólnie opisanej albo ograniczonej do wybranych etapów procesu. Ocena niska wskazuje natomiast, że publicznie dostępne dane nie pozwalają wiarygodnie potwierdzić istnienia wystarczająco rozwiniętej procedury kontroli. Taki sposób punktowania ma charakter porządkujący i porównawczy. Pozwala uchwycić różnice między domenami zastosowania GenAI, nie roszcząc sobie prawa do pełnej oceny skuteczności wdrożenia. Skala 0–3 w tabeli 3 ma charakter porządkujący i jakościowy. Nie jest niezależnym audytem wdrożeń, lecz transparentną rekonstrukcją dojrzałości mechanizmów opisanych w źródłach publicznych.

Tab. 3. Autorska punktacja nadzoru epistemicznego według ramy JWG-AI (skala 0–3)

Przypadek	Trafność	Dowodowość	Kompletność	Kontekst	Aktualność	Normatywność	Średnia
IBM AskHR	2,5	2,0	2,0	3,0	3,0	3,0	2,58
BMW Offer Analyst	2,5	3,0	3,0	2,5	2,0	2,5	2,58
Morgan Stanley Assistant/Debrief	3,0	3,0	2,5	3,0	3,0	3,0	2,92
ConsumerIQ ELC	2,5	2,5	2,0	2,5	3,0	2,0	2,42
McKinsey Lilli	2,5	3,0	2,5	3,0	2,5	3,0	2,75

Źródło: opracowanie własne na podstawie literatury i publicznych materiałów źródłowych wskazanych w bibliografii.

W syntetycznej ocenie porównawczej najwyżej oceniono przypadek Morgan Stanley. Wynik ten nie powinien być interpretowany jako dowód najwyższej efektywności ekonomicznej narzędzia. To po prostu konsekwencja najlepiej udokumentowanej (publicznie dostępnej) architektury kontroli jakości wiedzy – co więcej, w środowisku o podwyższonej wrażliwości regulacyjnej i decyzyjnej. O przewadze tego przypadku przesądza współwystępowanie kilku mechanizmów:

– ścisłego powiązania odpowiedzi ze źródłami wewnętrznymi,
– systematycznej ewaluacji jakości działania narzędzia,
– codziennych testów regresyjnych,
– podporządkowania procesu wymogom zgodności regulacyjnej oraz
– obowiązkowej weryfikacji treści przez doradcę przed ich użyciem w komunikacji z klientem.

Wysoki wynik McKinsey Lilli wynika z kolei z oparcia systemu na kuratorowanych zasobach wiedzy, obecności ról odpowiedzialnych za jakość i bezpieczeństwo danych, wskazywania źródeł oraz osadzenia narzędzia w profesjonalnym modelu pracy opartej na kapitale intelektualnym organizacji go reprezentującego. IBM AskHR i BMW Group Offer Analyst osiągają zbliżony poziom oceny, ale z odmiennych powodów. Otóż IBM wzmacnia jakość przez dostępność aktualnych polityk organizacyjnych, lokalne dopasowanie odpowiedzi i bezpieczeństwo normatywne procesów HR, natomiast

BMW – przez jawność kryteriów oceny oraz możliwość powiązania analizy zakupowej z dokumentacją źródłową. ConsumerIQ uzyskuje ocenę nieco niższą, ponieważ analityka wiedzy o rynku i konsumentach jest domeną mniej proceduralną, a bardziej interpretacyjną i narażoną na niejednoznaczność wnioskowania. Mimo że, co warte podkreślenia, narzędzie może znacząco skracać czas syntezy oraz wzmacniać aktualność rozpoznania trendów.

Ta komparatystyczna analiza prowadzi do wniosku, że jakość wiedzy generowanej przez GenAI nie podlega jednemu, uniwersalnemu wzorcowi oceny. Odpowiedź w obszarze HR nie jest tożsama z odpowiedzią w zakupach, doradztwie finansowym, marketingu czy konsultingu strategicznym. W HR szczególnego znaczenia nabiera zdolność systemu do rozpoznania sytuacji, w których automatyczna odpowiedź nie powinna zostać udzielona bez udziału człowieka. W zakupach zasadnicze stają się jawne kryteria porównania oraz możliwość wskazania fragmentów dokumentów, na których oparto ocenę danej oferty. W finansach jakość wymaga rekonstrukcji podstaw odpowiedzi, kontroli zgodności regulacyjnej i formalnej odpowiedzialności doradcy. Fundamentem w marketingu jest nie tylko szybkie porządkowanie rozproszonych danych, ale również odróżnienie trwałego trendu od szumu informacyjnego oraz wskazanie źródeł, z których wynika określony wniosek o rynku lub konsumencie. W doradztwie strategicznym system powinien nie tylko syntetyzować wiedzę. Ma ujawniać jej pochodzenie, ograniczenia, możliwe analogie oraz obszary, w których dostępny materiał nie pozwala na sformułowanie mocnej rekomendacji.

Z przeprowadzonej syntezy wyłaniają się zatem cztery mechanizmy o charakterze przekrojowym. Pierwszym jest merytoryczne porządkowanie posiadanego zasobu źródłowego. Wynika z niego pewna zasada – im większa waga decyzji, tym mniej wystarczające staje się korzystanie z otwartego, niekontrolowanego i niejednorodnego zasobu informacji. Drugim jest jawność podstaw źródłowych. Rozumie się przez to możliwość przejścia od wygenerowanej odpowiedzi do dokumentu, danych lub fragmentu korpusu wiedzy, na których oparto dane twierdzenie. Trzecim jest przełożenie zadania decyzyjnego na wyraźny układ kryteriów. Dzięki temu rekomendacja nie pozostaje ogólną sugestią modelu. Staje się tu wynikiem analizy prowadzonej według rozpoznawalnych i możliwych do zakwestionowania przesłanek. Czwartym mechanizmem jest ciągła ewaluacja działania systemu. Odpowiedzi GenAI powinny być testowane, porównywane, korygowane i ponownie oceniane w czasie. Jakość wiedzy generowanej przez AI nie jest jednorazową własnością modelu. Jest przecież rezultatem utrzymywanego procesu organizacyjnej kontroli.

Tab. 4. Przekrojowe mechanizmy kontroli jakości wiedzy generowanej przez systemy GenAI

Mechanizm kontroli jakości wiedzy	**Przypadki, w których mechanizm jest szczególnie widoczny**	**Znaczenie dla wiarygodności decyzji**	**Ryzyko w przypadku braku lub słabości mechanizmu**
Merytoryczne porządkowanie zasobu źródłowego	McKinsey Lilli, Morgan Stanley Debrief, IBM AskHR	Ogranicza materiał wykorzystywany przez system do dokumentów zatwierdzonych, aktualnych, opisanych i przydatnych w danej domenie organizacyjnej. Wzmacnia kontrolę nad tym, jakie źródła mogą stać się podstawą odpowiedzi.	Wykorzystanie treści nieaktualnych, nieuprawnionych albo niewłaściwych dla danej domeny, ryzyko naruszenia poufności, generowanie odpowiedzi niezgodnych z politykami organizacji.
Jawność podstaw źródłowych	Morgan Stanley Debrief, BMW Group Offer Analyst, ConsumerIQ, McKinsey Lilli	Pozwala powiązać twierdzenia systemu z dokumentami, danymi lub fragmentami repozytorium wiedzy. Umożliwia sprawdzenie odpowiedzi, odtworzenie jej podstaw oraz uzasadnienie decyzji wobec interesariuszy, klienta lub regulatora.	Odpowiedzi pozornie wiarygodne, lecz trudne do sprawdzenia, opóźnione wykrywanie błędów, brak możliwości obrony decyzji podjętej na podstawie rekomendacji systemu.
Przełożenie zadania na jawne kryteria oceny	BMW Group Offer Analyst, częściowo IBM AskHR i Morgan Stanley Debrief	Przekształca ogólne zapytanie w analizę prowadzoną według określonych warunków, wymagań lub kryteriów. Zwiększa porównywalność wariantów i ogranicza ryzyko pominięcia istotnej przesłanki.	Nadmiernie ogólna rekomendacja, arbitralność oceny, pominięcie warunku wykluczającego lub kryterium istotnego dla decyzji, fałszywe wrażenie kompletności analizy.
Ciągła ocena działania systemu i obieg informacji zwrotnej	Morgan Stanley Debrief, IBM AskHR	Umożliwia systematyczne sprawdzanie jakości odpowiedzi, identyfikowanie powtarzalnych błędów, porównywanie wyników w czasie oraz doskonalenie narzędzia na podstawie doświadczeń użytkowników.	Utrwalanie błędnych odpowiedzi, stopniowe pogarszanie jakości po zmianach modelu lub źródeł, brak wiedzy o granicach użyteczności systemu.

Mechanizm kontroli jakości wiedzy	Przypadki, w których mechanizm jest szczególnie widoczny	Znaczenie dla wiarygodności decyzji	Ryzyko w przypadku braku lub słabości mechanizmu
Przekazywanie przypadków nietypowych do oceny człowieka	IBM AskHR, Morgan Stanley Debrief	Chroni sprawy wymagające interpretacji eksperckiej, odpowiedzialności zawodowej lub oceny kontekstu, którego system nie powinien rozstrzygać samodzielnie. Wzmacnia rozróżnienie między sprawą rutynową a przypadkiem wrażliwym.	Automatyczne potraktowanie wyjątku jako sprawy standardowej, błędna decyzja w sytuacji wysokiej stawki, rozmycie odpowiedzialności za skutki odpowiedzi.
Oddzielenie syntezy wiedzy od decyzji menedżerskiej	McKinsey Lilli, ConsumerIQ, Morgan Stanley Debrief	Pozwala traktować system generatywnej sztucznej inteligencji jako narzędzie porządkowania i interpretowania materiału, a nie jako samodzielny podmiot decyzyjny. Utrzymuje odpowiedzialność po stronie użytkownika, eksperta lub właściwej funkcji organizacyjnej.	Podejmowanie decyzji na podstawie atrakcyjnej, lecz słabo udokumentowanej narracji, utożsamienie sprawnej syntezy z trafnością rekomendacji, przeniesienie odpowiedzialności na system.

Źródło: opracowanie własne na podstawie analizy przypadków IBM AskHR, BMW Group Offer Analyst, AI @ Morgan Stanley Assistant/Debrief, ConsumerIQ oraz McKinsey Lilli.

Od *human-in-the-loop*[12] do *epistemic accountability*[13]

Najważniejszy wniosek z tego badania jest następujący. Jakość wiedzy generowanej przez AI jest właściwością układu socjotechnicznego, a nie samego modelu. Model może zwiększać płynność, zakres syntezy i szybkość dostępu do treści, ale jakość decyzyjna pojawia się dopiero wtedy, kiedy wynik zostaje powiązany z korpusem, kryteriami, kontekstem,

[12] *Human-in-the-loop* (ang.) oznacza udział człowieka w pętli decyzyjnej, tj. taki sposób organizacji pracy z systemem AI, w którym wynik modelu podlega weryfikacji przez osobę posiadającą kompetencje, uprawnienie do interwencji i odpowiedzialność za jego użycie. Wytyczne Komisji Europejskiej [*Ethics Guidelines for Trustworthy AI*, 2019; art. 14 Regulation (EU) 2024/1689] dotyczące godnej zaufania AI wskazują, że nadzór może przyjmować postać *human-in-the-loop*, *human-on-the-loop* lub *human-in-command*, a AI Act wymaga, aby człowiek mógł rozumieć ograniczenia systemu, monitorować jego działanie, przeciwdziałać nadmiernemu zaufaniu do wyniku, interpretować odpowiedź, odrzucić ją lub zatrzymać działanie systemu.

[13] *Epistemic accountability* (ang.) to odpowiedzialność za wiarygodność wiedzy oznaczająca obowiązek uzasadnienia, dlaczego dane twierdzenie, rekomendacja lub synteza mogą zostać uznane za wiarygodną podstawę decyzji.

aktualnością i, wreszcie, odpowiedzialnością. To przesuwa dyskusję z poziomu „jak dobry jest model" na poziom „jak dobrze organizacja projektuje warunki użycia tego modelu".

Popularna formuła *human-in-the-loop* jest w tym kontekście zbyt uboga. Nie mówi, który człowiek ma być w pętli, na jakim etapie, z jakimi kompetencjami i wobec jakiego ryzyka. W analizowanych przypadkach można wyróżnić co najmniej pięć ról ludzkich. Pierwsza to właściciel korpusu, odpowiedzialny za jakość źródeł. Druga to projektant kryteriów, definiujący warunki oceny. Trzecia to użytkownik końcowy, który formułuje zapytanie i interpretuje odpowiedź. Czwarta to ekspert walidujący, który sprawdza trafność w zadaniach wysokiej stawki. Piąta to funkcja *governance/compliance*, która określa granice użycia, testy i procedury dalszej eskalacji.

Jeżeli te role zostaną sprowadzone do jednej ogólnej odpowiedzialności użytkownika, organizacja tworzy ryzyko moralnej i epistemicznej próżni. Użytkownik końcowy może nie mieć ani wiedzy o jakości korpusu, ani możliwości oceny tak zwanej metody *retrieval*, ani kompetencji prawnej do oceny skutków normatywnych. Stąd konieczność przejścia od *human-in-the-loop* do *epistemic accountability*, czyli jasnego przypisania odpowiedzialności za każdy etap produkcji i użycia wiedzy generowanej przez AI.

Wyniki studium prowadzą do pięciu propozycji teoretycznych. Można je ująć w następujący sposób:

- w zadaniach wysokiej stawki możliwość śledzenia dowodów będzie wyraźniej powiązana z użytecznością decyzyjną niż płynność językowa odpowiedzi,
- GenAI zwiększy jakość decyzji szczególnie wtedy, gdy zadanie zostanie przełożone na jawne kryteria oceny,
- w organizacjach regulowanych rutyny ewaluacyjne będą warunkiem adopcji, a nie jedynie narzędziem optymalizacji,
- skuteczność GenAI będzie zależała od zdolności organizacji do rozpoznawania granicy kompetencji systemu,
- narzędzia oparte na silnym korpusie wewnętrznym zwiększą spójność wiedzy, ale mogą także wzmacniać epistemiczny konserwatyzm.

Dla nauk o jakości szczególnie istotna jest propozycja czwarta. Jakość w środowisku GenAI nie jest stanem statycznym. Jest dynamicznym procesem kontroli błędu. Obejmuje projektowanie wejścia, ograniczenie przestrzeni generowania, widoczność źródeł, audyt odpowiedzi, feedback, aktualizację korpusu i uczenie użytkownika. To zbliża analizę GenAI do klasycznych problemów zarządzania jakością, czyli kontroli procesu, zapobiegania defektom, identyfikacji przyczyn źródłowych i ciągłego doskonalenia. Różnica polega na tym, że „defektem" nie jest wyłącznie wadliwy produkt, lecz wadliwa przesłanka decyzji.

Warto też zauważyć, że GenAI zmienia ekonomię wiedzy w organizacji. Dotychczas kosztem była często trudność dotarcia do wiedzy. Po wdrożeniu GenAI koszt przesuwa się w stronę oceny jakości wiedzy. Gdy dostęp jest łatwy, największym zagrożeniem staje się nie brak informacji, ale nadmiar przekonująco sformułowanych przesłanek. Organizacje dojrzałe będą więc różnić się zakresem adopcji AI. Różnicą tą będzie również zdolność do odsiewania odpowiedzi niepewnych, nieadekwatnych i nieudokumentowanych.

Implikacje praktyczne dla menedżerów i projektantów systemów jakości

Z praktycznego punktu widzenia artykuł prowadzi do wniosku, że wdrożenie GenAI powinno zaczynać się od mapy ryzyka decyzyjnego, a nie od wyboru narzędzia. Organizacja powinna najpierw ustalić, jakie szkody może spowodować błędna odpowiedź. Mogą to być szkody prawne, finansowe, reputacyjne, operacyjne, relacyjne albo poznawcze. Dopiero potem należy rozpocząć projektowanie poziomu dostępu do źródeł, widoczności cytowań, zakresu automatyzacji, eskalacji i obowiązków użytkownika.

W HR rekomendowane są: ograniczone domeny pytań, lokalizacja polityk, kontrola wersji, ścieżki eskalacji dla wyjątków i regularna analiza feedbacku. W zakupach istotne są natomiast: jawne kryteria oceny, macierze porównawcze, linkowanie odpowiedzi do fragmentów ofert i rozdzielenie analizy formalnej od decyzji strategicznej. W sektorach regulowanych, zwłaszcza w finansach, wiarygodność wiedzy generowanej przez GenAI zależy od zespolenia kilku warstw kontroli: cyklicznej oceny jakości odpowiedzi, testów regresyjnych, rejestrowania użycia systemu, ograniczenia przechowywania danych, weryfikacji eksperckiej oraz jednoznacznego informowania użytkownika o ograniczeniach narzędzia. W marketingu problem jakości ma inny charakter, ponieważ dotyczy nie tyle zgodności z ustaloną procedurą, ile rzetelności interpretacji rozproszonych sygnałów rynkowych. W tym przypadku konieczna jest kontrola różnorodności źródeł, dat ich powstania, lokalnego kontekstu danych oraz stopnia potwierdzenia formułowanych wniosków o konsumentach. W doradztwie strategicznym pierwszorzędne staje się natomiast instytucjonalne zarządzanie zasobami wiedzy. Ich selekcja, opis, aktualizacja, ochrona poufności i ocena przydatności to obszary dające temu nowy sens i jakość. GenAI nie powinna w tej domenie jedynie przywoływać podobnych przypadków z przeszłości. Przede wszystkim ma wskazywać, gdzie analogia jest niepełna, gdzie materiał źródłowy pozostaje zbyt ubogi, a gdzie rekomendacja wymaga dalszej oceny eksperckiej.

Projektowanie jakości wiedzy GenAI można ująć w pięciu krokach. Pierwszy krok to klasyfikacja zadań według stawki błędu. Drugi to przypisanie minimalnych wymogów dowodowych dla każdej klasy. Trzeci to zdefiniowanie kryteriów i checklist, które model ma obsługiwać. Czwarty to budowa pętli ewaluacyjnej: testy, regresja, feedback, monitoring jakości i korekty *retrieval*. Piąty to szkolenie użytkowników nie tylko z promptowania, ale przede wszystkim z kalibracji zaufania i rozpoznawania granic modelu.

Zarządzanie jakością GenAI nie może kończyć się na polityce użycia. Polityka ta jest konieczna, ale niewystarczająca. Potrzebne są tu konkretne artefakty jakościowe – od rejestru źródeł, karty ryzyka przypadku użycia, zestawu testów referencyjnych, checklisty walidacji informacji wychodzącej, katalogu przypadków eskalacji, metryki niezgodności, po procedurę wycofania lub korekty odpowiedzi. Dopiero taki zestaw praktyk zmienia GenAI z nieformalnego asystenta w element systemu zarządzania wiedzą i jakością – w służbie różnych branż i specjalizacji.

Ograniczenia badania

Badanie ma trzy główne ograniczenia. Po pierwsze, opiera się na publicznych materiałach organizacji i partnerów technologicznych. Źródła te pozwalają zrekonstruować architekturę wdrożeń, ale nie dają pełnego dostępu do danych operacyjnych, nieudanych odpowiedzi, incydentów jakościowych i wewnętrznych kosztów *governance*. Po drugie, punktacja JWG-AI ma charakter autorskiej oceny jakościowej. Jest przejrzysta, ale nie zastępuje niezależnego audytu ani pomiaru międzykoderowego. Po trzecie, przypadki pochodzą głównie z dużych organizacji o wysokich zasobach technologicznych i kompetencyjnych, dlatego nie można automatycznie przenosić wniosków na MŚP.

Ograniczenia te nie unieważniają wniosków, bardziej określają ich status. Artykuł nie dowodzi, że opisane wdrożenia poprawiły wszystkie decyzje menedżerskie. Pokazuje natomiast, jakie mechanizmy jakości są widoczne w najbardziej dojrzałych publicznie opisanych zastosowaniach i jak można je konceptualizować w języku nauk o zarządzaniu i jakości.

Podsumowanie

W epoce GenAI zasadniczym problemem zarządzania nie jest już samo wytwarzanie odpowiedzi. Problemem staje się rozstrzygnięcie, kiedy odpowiedź może stać się wiedzą, a wiedza – odpowiedzialną przesłanką decyzji. Przeprowadzone porównawcze studium przypadków pokazuje, że organizacje osiągają wyższą jakość nie przez nieograniczone oddanie decyzji modelowi, ale przez różne ograniczenia.

GenAI może wzmacniać osąd menedżerski, gdy skraca drogę do właściwych źródeł, porządkuje kryteria i ujawnia uzasadnienie. Może jednak również produkować pozór wiedzy – odpowiedzi eleganckie, szybkie i przekonujące. Epistemicznie będą jednak słabe. Wysoka jakość nie polega więc na tym, że model mówi pewniej. Polega na tym, że organizacja projektuje warunki, w których użytkownik może mniej naiwnie ufać, lepiej sprawdzać i odpowiedzialniej decydować.

Z perspektywy nauk o zarządzaniu i jakości proponowana rama JWG-AI pozwala przesunąć debatę od produktywności do wiarygodności. Wzrost szybkości jest ważny, ale nie stanowi samodzielnej miary wartości GenAI. W decyzjach menedżerskich najważniejsze pytanie brzmi, jaki jest koszt błędu, którego dzięki AI nie zauważono szybciej. Odpowiedź na to pytanie powinna wyznaczać przyszłe badania, projekty eksperymentalne i praktykę zarządzania systemami generatywnej sztucznej inteligencji.

Bibliografia

Agrawal A., Gans J., Goldfarb A. (2018), *Prediction Machines. The Simple Economics of Artificial Intelligence*, Boston: Harvard Business Review Press.

Alasuutari P., Qadir A. (2014), *Epistemic governance: an approach to the politics of policy-making*, „European Journal of Cultural and Political Sociology", vol. 1(1), s. 67–84, https://doi.org/10.1080/23254823.2014.887986

Alavi M., Leidner D.E. (2001), *Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues*, „MIS Quarterly", vol. 25(1), s. 107–136.

Artificial Intelligence Risk Management Framework (AI RMF 1.0) (2023), https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10 [dostęp: 15.05.2026].

Brynjolfsson E., Li D., Raymond L.R. (2025), *Generative AI at Work*, „The Quarterly Journal of Economics", vol. 140(2), s. 889–942, https://doi.org/10.1093/qje/qjae044

Csaszar F.A., Ketkar H., Kim H. (2024), *Artificial Intelligence and Strategic Decision-Making: Evidence from Entrepreneurs and Investors*, „Strategy Science", vol. 9(4), s. 322–345, https://doi.org/10.1287/stsc.2024.0190

Davenport T.H., Prusak L. (1998), *Working Knowledge. How Organizations Manage What They Know*, Boston: Harvard Business School Press.

Dell'Acqua F., McFowland III E., Mollick E., Lifshitz H., Kellogg K.C., Rajendran S., Krayer L., Candelon F., Lakhani K.R. (2026), *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality*, „Organization Science", vol. 37(2), s. 403–423, https://doi.org/10.1287/orsc.2025.21838

Eisenhardt K.M. (1989), *Building theories from case study research*, „Academy of Management Review", vol. 14(4), s. 532–550, https://doi.org/10.5465/amr.1989.4308385

Eisenhardt K.M., Graebner M.E. (2007), *Theory building from cases: Opportunities and challenges*, „Academy of Management Journal", vol. 50(1), s. 25–32, https://doi.org/10.5465/amj.2007.24160888

ELC Partners With Microsoft to Create ConsumerIQ Tool for Data-Driven Insights (2025), https://www.elcompanies.com/en/news-and-media/newsroom/company-features/2025/elc-microsoft-consumer-iq [dostęp: 15.05.2026].

Ethics Guidelines for Trustworthy AI (2019), High-Level Expert Group on Artificial Intelligence, Brussels: European Commission.

Faraj S., Pachidi S., Sayegh K. (2018), *Working and organizing in the age of the learning algorithm*, „Information and Organization", vol. 28(1), s. 62–70, https://doi.org/10.1016/j.infoandorg.2018.02.005

Gibbert M., Ruigrok W., Wicki B. (2008), *What passes as a rigorous case study?*, „Strategic Management Journal", vol. 29(13), s. 1465–1474, https://doi.org/10.1002/smj.722

Glikson E., Woolley A.W. (2020), *Human trust in artificial intelligence: Review of empirical research*, „Academy of Management Annals", vol. 14(2), s. 627–660, https://doi.org/10.5465/annals.2018.0057

Greater efficiency and productivity with artificial intelligence - generative AI in BMW Group Purchasing (2025), https://www.press.bmwgroup.com/global/article/detail/T0450032EN/ [dostęp: 15.05.2026].

ISO/IEC 42001:2023. Information technology – Artificial intelligence – Management system (2023), https://www.iso.org/standard/42001 [dostęp: 15.05.2026].

Jarrahi M.H. (2018), *Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making*, „Business Horizons", vol. 61(4), s. 577–586, https://doi.org/10.1016/j.bushor.2018.03.007

Lee J.D., See K.A. (2004), *Trust in automation: Designing for appropriate reliance*, „Human Factors", vol. 46(1), s. 50–80.

Lee Y.W., Strong D.M., Kahn B.K., Wang R.Y. (2002), *AIMQ: A methodology for information quality assessment*, „Information & Management", vol. 40(2), s. 133–146, https://doi.org/10.1016/S0378-7206(02)00043-5

Meet Lilli, our generative AI tool that's a researcher, a time saver, and an inspiration (2023), https://www.mckinsey.com/about-us/new-at-mckinsey-blog/meet-lilli-our-generative-ai-tool [dostęp: 15.05.2026].

Morgan Stanley Research Announces AskResearchGPT Powered by OpenAI (2024), https://www.morganstanley.com/press-releases/morgan-stanley-research-announces-askresearchgpt [dostęp: 15.05.2026].

Morgan Stanley uses AI evals to shape the future of financial services (2025), https://openai.com/index/morgan-stanley/ [dostęp: 15.05.2026].

Morgan Stanley Wealth Management Announces Key Milestone in Innovation Journey with OpenAI (2023), https://www.morganstanley.com/press-releases/key-milestone-in-innovation-journey-with-openai [dostęp: 15.05.2026].

Morgan Stanley Wealth Management Announces Latest Game-Changing Addition to Suite of GenAI Tools (2024), https://www.morganstanley.com/press-releases/ai-at-morgan-stanley-debrief-launch [dostęp: 15.05.2026].

Nonaka I. (1994), *A dynamic theory of organizational knowledge creation*, „Organization Science", vol. 5(1), s. 14–37, https://doi.org/10.1287/orsc.5.1.14

Noy S., Zhang W. (2023), *Experimental evidence on the productivity effects of generative artificial intelligence*, „Science", vol. 381(6654), s. 187–192, https://doi.org/10.1126/science.adh2586

OECD AI Principles overview (2024), https://oecd.ai/en/ai-principles [dostęp: 15.05.2026].

Otis N.G., Clarke R., Delecourt S., Holtz D., Koning R. (2024), *The Uneven Impact of Generative AI on Entrepreneurial Performance*, https://www.hbs.edu/ris/Publication%20Files/24-042_567db548-5b24-4182-b703-9d5ef1328043.pdf [dostęp: 15.05.2026].

Raisch S., Krakowski S. (2021), *Artificial intelligence and management: The automation-augmentation paradox*, „Academy of Management Review", vol. 46(1), s. 192–210, https://doi.org/10.5465/amr.2018.0072

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

Revamping procurement operations with generative AI-powered Offer Analyst solution on AWS (2024), https://aws.amazon.com/blogs/industries/revamping-procurement-operations-with-generative-ai/ [dostęp: 15.05.2026].

Shrestha Y.R., Ben-Menahem S.M., von Krogh G. (2019), *Organizational decision-making structures in the age of artificial intelligence*, „California Management Review", vol. 61(4), s. 66–83, https://doi.org/10.1177/0008125619862257

Simon H.A. (1979), *Rational decision making in business organizations*, „American Economic Review", vol. 69(4), s. 493–513, https://www.jstor.org/stable/1808698

Sperber D., Clément F., Heintz C., Mascaro O., Mercier H., Origgi G., Wilson D. (2010), *Epistemic Vigilance*, „Mind & Language", vol. 25(4), s. 359–393, https://doi.org/10.1111/j.1468-0017.2010.01394.x

The Estée Lauder Companies and Microsoft Copilot Studio transform intelligence (2025), https://www.microsoft.com/en/customers/story/23488-the-estee-lauder-companies-microsoft-copilot-studio [dostęp: 15.05.2026].

Transforming HR support with agentic AI (2025), https://www.ibm.com/case-studies/ibm-askhr [dostęp: 15.05.2026].

Wand Y., Wang R.Y. (1996), *Anchoring data quality dimensions in ontological foundations*, „Communications of the ACM", vol. 39(11), s. 86–95, https://doi.org/10.1145/240455.240479

Wang R.Y., Strong D.M. (1996), *Beyond accuracy: What data quality means to data consumers*, „Journal of Management Information Systems", vol. 12(4), s. 5–33, https://doi.org/10.1080/07421222.1996.11518099

Weick K.E. (1995), *Sensemaking in Organizations*, Thousand Oaks: Sage.

Yin R.K. (2018), *Case Study Research and Applications: Design and Methods*, 6th ed., Thousand Oaks: Sage.

O autorze

Bartosz Rzętkiewicz – mgr, wicekanclerz ds. rozwoju Wyższej Szkoły Biznesu i Nauk o Zdrowiu w Łodzi, dyrektor Instytutu Prawa i Kryminologii oraz Instytutu Bezpieczeństwa Wewnętrznego Wyższej Szkoły Biznesu i Nauk o Zdrowiu w Łodzi.

About the Author

Bartosz Rzętkiewicz – MA, Deputy Chancellor for Development at the Academy of Business and Health Sciences in Łódź, and Director of the Institute of Law and Criminology and the Institute of Internal Security.

Ten utwór jest dostępny na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe.

Open Access