Delegowanie osądu i erozja sprawczości w relacji człowiek-AI w świetle książki The AI Instinct Rany Gujral

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje przejście od automatyzacji prostych zadań do delegowania osądu na systemy AI, co prowadzi do erozji ludzkiej sprawczości. Autor wskazuje na zjawisko cognitive offloading i ryzyko deskillingu – utraty kluczowych kompetencji poznawczych w wyniku rzadszego ich samodzielnego stosowania. Problem nie wynika z lenistwa, lecz z adaptacji do nowej struktury kosztów poznawczych, gdzie krótkoterminowa efektywność koliduje z długofalowym kapitałem intelektualnym. Szczególną uwagę zwrócono na paradoks kompetencji: podczas gdy obecni eksperci traktują AI jako rozszerzenie wiedzy, nowe pokolenia mogą tracić zdolność samodzielnego nadzoru nad technologią, co staje się krytycznym wyzwaniem w obszarach takich jak medycyna.

The article analyzes the transition from automating simple tasks to delegating judgment to AI systems, which leads to the erosion of human agency. The author points to the phenomenon of cognitive offloading and the risk of deskilling—the loss of key cognitive competencies as a result of using them independently less frequently. This problem does not stem from laziness, but from adaptation to a new structure of cognitive costs, where short-term efficiency conflicts with long-term intellectual capital. Particular attention is paid to the competency paradox: while current experts treat AI as an extension of knowledge, new generations may lose the ability to independently oversee technology, which becomes a critical challenge in fields such as medicine.

Artykuł analizuje krytyczne zagrożenie wynikające z ewolucji relacji człowiek-AI, wskazując, że prawdziwym problemem nie jest zastąpienie ludzkiej pracy przez automatyzację, lecz subtelne delegowanie procesów poznawczych i osądu na systemy

algorytmiczne. Autor stawia tezę, że bezrefleksyjne korzystanie z AI prowadzi do deskillingu – utraty kluczowych kompetencji analitycznych – oraz tworzy iluzję sprawczości, w której człowiek staje się jedynie formalnym zatwierdzaczem decyzji wygenerowanych przez maszynę. Tekst argumentuje, że aby uniknąć poznawczej homogenizacji i pułapki 'moralnej strefy zgniotu', należy odejść od pasywnego modelu nadzoru (human in the loop) na rzecz aktywnej sprawczości (human agency in the loop). Kluczem do zachowania niezależności intelektualnej ma być projektowanie systemów opartych na 'proaktywnym tarciu' – celowych utrudnieniach poznawczych, które wymuszają krytyczną analizę i chronią zdolność człowieka do samodzielnego konstruowania przestrzeni wyboru.

SŁOWA KLUCZOWE

delegowanie osądu | erozja sprawczości | cognitive offloading | automation bias | deskilling | proaktywne tarcie | cognitive forcing | epistemic dependence | meaningful human control | moral crumple zone | pluralizm epistemiczny | human agency in the loop | hybrydowy system poznawczy | koewolucja kognitywna | interakcyjna epistemologia

Delegowanie osądu prowadzi do deskillingu

Od delegowania zadań do delegowania osądu: cognitive offloading, automation bias i przebudowa ludzkiej sprawczości

Kluczowym problemem w relacji człowieka ze sztuczną inteligencją nie jest sama automatyzacja pracy, lecz systematyczne przesuwanie granicy między tym, co wykonujemy osobiście, a tym, co pozostaje integralnym elementem naszego procesu poznawczego. Można bowiem powierzyć maszynie rachunki, pozostając autorem problemu, kryterium rozwiązania i końcowej oceny wyniku. Można zlecić jej wyszukanie informacji, zachowując samodzielność interpretacji. Istnieje jednak ryzyko przekazania AI również określania istotności danych, generowania hipotez, prognozowania ryzyka, hierarchizacji alternatyw czy rekomendowania działań – a w konsekwencji przygotowania uzasadnienia, które pozwoli przedstawić decyzję jako racjonalną. W takim scenariuszu człowiek formalnie nadal pozostaje osobą „podejmującą decyzję", lecz coraz większa

część struktury poznawczej poprzedzającej wybór znajduje się poza nim. To właśnie tutaj przebiega granica pomiędzy automatyzacją czynności a delegowaniem osądu.

Psychologia poznawcza od dawna analizuje zjawisko cognitive offloading, czyli przenoszenia części operacji poznawczych na zewnętrzne artefakty lub środowisko. Zapisujemy termin spotkania, by nie obciążać pamięci; korzystamy z mapy, zamiast rekonstruować trasę; używamy kalkulatora, aby nie przechowywać wyników pośrednich w pamięci roboczej. Takie odciążenie nie jest aberracją, lecz fundamentalną cechą ludzkiej inteligencji. Rozszerzone poznanie rozwija się właśnie dzięki temu, że ograniczony biologicznie system potrafi modyfikować otoczenie tak, aby zredukować koszt własnych operacji.

Problem pojawia się w momencie, gdy narzędzie przejmuje nie tylko obciążenie pamięciowe czy obliczeniowe, ale funkcje niezbędne do rozpoznania, czy dany wynik zasługuje na akceptację. Generatywna AI zasadniczo zmienia skalę tego zjawiska, ponieważ system zewnętrzny nie ogranicza się już do przechowywania dostarczonych treści. Potrafi tworzyć własną syntezę, ustalać strukturę dokumentu, dobierać argumenty, sugerować priorytety czy redagować komunikaty. System przesuwa się zatem w górę hierarchii poznawczej: od pamięci pomocniczej ku zewnętrznemu generatorowi reprezentacji problemu.

To przesunięcie jest istotniejsze niż sam wzrost produktywności, gdyż decyzja człowieka zależy od tego, jak wcześniej została skonstruowana przestrzeń alternatyw. Książka opisuje ten proces poprzez metaforę "duchów w pętli sprzężenia zwrotnego" oraz ostrzeżenie przed nadmiernie gładką współpracą z AI. Gujral zauważa, że system może wpływać na ludzkie poznanie poprzez subtelną modulację procesów poprzedzających refleksyjny osąd, a brak "tarcia" poznawczego może prowadzić do sytuacji, w której człowiek zaczyna raczej konsumować gotową myśl niż samodzielnie angażować się w jej wytwarzanie.

W ujęciu naukowym należy odrzucić zbyt prostą wizję, według której AI stopniowo „wyłącza System 2”. Mechanizm ten jest bardziej złożony: technologia zmienia koszt poszczególnych strategii poznawczych, a człowiek – zgodnie z zasadami ograniczonej racjonalności – adaptuje się do nowej struktury tych kosztów. Jeśli samodzielne przeszukiwanie źródeł zajmuje godzinę, a otrzymanie syntezy kilka sekund, korzystanie z tej drugiej opcji jest racjonalne. Podobnie jest w przypadku przygotowania pierwszej wersji argumentacji.

Problem nie wynika zatem z pierwotnego "lenistwa poznawczego". Pojawia się on wtedy, gdy krótkoterminowo racjonalna strategia oszczędzania wysiłku zmienia długofalową strukturę kompetencji. Jednostka może systematycznie wybierać optymalną lokalnie ścieżkę współpracy z AI, by po kilku latach odkryć, że posiada mniejszą zdolność samodzielnego wykonywania operacji

niezbędnych do nadzoru nad tą samą technologią. Jest to klasyczny konflikt między efektywnością krótkookresową a kapitałem poznawczym.

Zjawisko to najlepiej opisuje pojęcie deskilling. Nie oznacza ono generalnego spadku inteligencji, lecz utratę lub osłabienie konkretnej kompetencji wskutek rzadszego jej samodzielnego stosowania. Przegląd dotyczący wykorzystania AI w medycynie z 2026 roku wskazał już empiryczne przesłanki tego ryzyka: obserwowano pogorszenie samodzielności po pracy z systemem AI, podatność diagnostów na błędne podpowiedzi oraz ograniczenie liczby przypadków, na których młodzi profesjonaliści mogą ćwiczyć diagnozę. Autorzy podkreślali, że choć dowody są wciąż ograniczone, są one wystarczająco spójne, by uznać ochronę kompetencji za element bezpieczeństwa klinicznego.

Kwestia ta odsłania szczególnie trudny paradoks kompetencji. Dzisiejszy ekspert zazwyczaj uczył się zawodu w świecie, w którym AI odgrywała mniejszą rolę; posiada więc samodzielny model dziedziny i traktuje technologię jako rozszerzenie wiedzy. Następne pokolenie może jednak zdobywać kompetencje w środowisku, gdzie część zadań jest automatyczna od samego początku. Badanie jakościowe z 2025 roku ukazało ten rozdźwięk: doświadczeni pracownicy ochrony zdrowia nie odczuwali silnego deskillingu, lecz obawiali się, że młodsi klinicyści mogą nie mieć wystarczającej liczby okazji do rozwinięcia niezależnego osądu, ponieważ część pracy diagnostycznej przejmie AI.

Nie jest to zjawisko wyjątkowe dla medycyny.

Augmentacja zamiast automatyzacji chroni kompetencje poznawcze

W programowaniu generowanie kodu może ograniczać liczbę sytuacji, w których początkujący programista musi samodzielnie przejść od problemu do struktury rozwiązania. W prawie automatyczne streszczenia orzecznictwa mogą skracać proces rozwijania umiejętności krytycznej analizy i porównywania źródeł. W administracji model tworzący projekt decyzji może rzadziej zmuszać urzędnika do samodzielnej rekonstrukcji podstawy prawnej, a w nauce automatyczna synteza literatury może ograniczać ekspozycję na sprzeczności i słabe sygnały obecne w źródłach pierwotnych. Efekt końcowy zależy od sposobu użycia: ta sama technologia może przyspieszać rozwój eksperckości, jeśli pełni rolę nauczyciela i partnera krytycznego, albo ją osłabiać, jeśli konsekwentnie eliminuje poznawczo formacyjne etapy pracy. Dlatego opozycja augmentation versus automation jest ważniejsza niż proste pytanie o to, czy AI została "użyta". Augmentacja zwiększa możliwości człowieka, zachowując lub rozwijając kompetencje niezbędne do sprawowania

kontroli nad wynikiem. Automatyzacja natomiast może optymalizować sam rezultat, niezależnie od rozwoju operatora. W krótkim okresie oba modele mogą produkować identycznie dobrą decyzję; różnica między nimi ujawnia się jednak w czasie, gdy system przestaje działać, pojawia się sytuacja nietypowa lub algorytm popełnia błąd.

Organizacja oparta na augmentacji zachowuje rezerwę ludzkich kompetencji. Organizacja oparta na głębokiej substytucji może odkryć, że formalnie obecny człowiek nie potrafi już samodzielnie odtworzyć procesu, który ma nadzorować. Na tym tle należy analizować automation bias – tendencję do nadmiernego polegania na rekomendacjach systemów zautomatyzowanych, co objawia się przyjmowaniem błędnych sugestii lub pomijaniem informacji, których system nie wskazał. Najnowszy przegląd 35 badań dotyczących human-AI collaboration pokazuje, że automation bias nie wynika z jednego mechanizmu. Wpływają na niego między innymi poziom wiedzy użytkownika, AI literacy, charakter zadania, złożoność wyjaśnienia, dynamika zaufania, koszt weryfikacji oraz indywidualne różnice poznawcze. Właściwym problemem nie jest zatem samo "zaufanie do AI", lecz kalibracja zaufania względem rzeczywistej wiarygodności systemu w konkretnej klasie sytuacji. Błędem jest również zbyt małe zaufanie – jeśli system jest niezawodnie lepszy od człowieka w określonej domenie, uporczywe odrzucanie jego rekomendacji obniża jakość decyzji. Istnieje zatem zarówno automation bias, jak i algorithm aversion.

Celem dobrze zaprojektowanej współpracy nie jest maksymalizowanie ludzkiej niezależności, lecz uzyskanie racjonalnej dystrybucji epistemicznego autorytetu. Człowiek powinien wiedzieć, w jakich sytuacjach system jest statystycznie lepszy, gdzie zawodzi, jakie informacje wykorzystuje i jaką wagę nadać jego rekomendacji. Jest to problem zaufania warunkowego, a nie ogólnego. Najbardziej pouczające są przypadki, w których człowiek posiada prawidłowy sąd początkowy, a następnie zmienia go pod wpływem błędnej sugestii AI. W badaniach klinicznych od dawna obserwuje się takie odwrócenia; nowsze dane potwierdzają, że błędna rekomendacja może istotnie pogarszać trafność decyzji. Badanie opublikowane w 2026 roku wykazało, że poprawne rekomendacje AI podnosiły wyniki diagnostyczne, natomiast błędne wyraźnie je obniżały – kluczowym czynnikiem ochronnym pozostawała własna kompetencja profesjonalisty. Oznacza to, że system hybrydowy nie może być oceniany poprzez niezależną dokładność człowieka i modelu. Interakcja tworzy trzeci układ poznawczy, którego właściwości nie są prostą sumą komponentów. Możemy to zapisać jako problem interakcyjnej epistemologii. Jeżeli człowiek ma rację w 80 procentach przypadków, a AI w 90 procentach, nie wynika z tego automatycznie, że zespół człowiek-AI osiągnie wynik powyżej 90 procent.

Wysoki wynik jest możliwy tylko wtedy, gdy człowiek potrafi rozpoznać momenty, w których model się myli, i vice versa. Z drugiej strony, rezultat może być gorszy od obu elementów, jeśli

rekomendacje systemu destabilizują trafne sądy człowieka lub człowiek ingeruje w prawidłowe wyniki modelu. Jakość współpracy zależy więc przede wszystkim od komplementarności błędów oraz mechanizmu arbitrażu.

Proaktywne tarcie chroni przed bezmyślnym uleganiem rekomendacjom

W tym miejscu pojawia się pojęcie cognitive forcing – celowego projektowania interfejsu tak, aby spowolnić automatyczne przyjęcie rekomendacji i wymusić niezależne przetwarzanie informacji. Można na przykład poprosić użytkownika o wstępną ocenę przed wyświetleniem sugestii AI, opóźnić prezentację wyniku lub wymagać wskazania argumentu przemawiającego przeciwko propozycji systemu. Gujral proponuje podobny mechanizm pod nazwą "proaktywnego tarcia", wiążąc go z przedstawianiem alternatyw, ujawnianiem niepewności, możliwością wycofania działania oraz obowiązkiem autoryzacji decyzji o wysokiej stawce. Jednocześnie najnowsze badania wskazują, że samo spowolnienie procesu nie jest panaceum.

W eksperymentach dotyczących oceny ryzyka przemocy w sytuacjach kryzysu psychiatrycznego stronnicze rekomendacje AI prowadziły do bardziej stronniczych decyzji uczestników; zastosowane techniki cognitive forcing nie zniwelowały tego efektu. Większą odporność wykazywały osoby skłonne do angażowania się w wysiłkowe przetwarzanie poznawcze, choć mechanizm ten zależy od poziomu zmęczenia i kontekstu. Wniosek jest istotny: tarcie poznawcze działa tylko wtedy, gdy prowadzi do realnej analizy alternatywy, a nie staje się kolejnym formalnym kliknięciem. Nie wystarczy zatem dodać przycisku "czy jesteś pewwna?". Użytkownik szybko nauczy się klikać go automatycznie, podobnie jak akceptuje regulaminy czy ostrzeżenia bezpieczeństwa. Projektowanie autonomii wymaga stworzenia sytuacji, w której informacja niezbędna do krytycznej oceny rzeczywiście staje się poznawczo dostępna. Czasami lepszym rozwiązaniem będzie pokazanie argumentów przeciw rekomendacji, przedstawienie kilku konkurujących hipotez, ujawnienie, że przypadek wykracza poza typowy rozkład treningowy, a niekiedy całkowite powstrzymanie się AI od udzielenia jednej odpowiedzi. Kluczowa jest tutaj kolejność informacji.

Jeśli człowiek najpierw otrzymuje stanowisko AI, jego późniejsza analiza nie rozpoczyna się w stanie neutralnym. Rekomendacja staje się kotwicą, wpływa na interpretację dowodów i ustanawia domyślną hipotezę. Dlatego rozwiązanie, w którym profesjonalista dokonuje niezależnej oceny przed ujawnieniem wyniku modelu, może chronić epistemiczną różnorodność zespołu. Nie jest to zawsze praktyczne ani efektywne, ale pokazuje, że jeśli chcemy uzyskać komplementarność dwóch systemów poznawczych, nie powinniśmy doprowadzać do sytuacji, w której drugi system

natychmiast podporządkowuje pierwszy własnemu punktowi wyjścia. Problem ustawień domyślnych wykracza daleko poza decyzje eksperckie. Psychologia wyboru od dziesięcioleci wskazuje na znaczenie default effects: opcja wstępnie wybrana zyskuje przewagę wynikającą z bezwładu, kosztu zmiany, interpretacji rekomendacji i ograniczonej uwagi. W systemach AI ustawieniem domyślnym może być nie tylko zaznaczone pole, ale pierwsza wygenerowana odpowiedź. Jeśli system prezentuje jedną propozycję jako naturalny rezultat analizy, alternatywy zyskują status czegoś wymagającego dodatkowego wysiłku. Badania z 2026 roku dotyczące algorytmicznych ustawień domyślnych w decyzjach społecznych wykazały, że ekspozycja na predefiniowane preferencje algorytmu może wpływać na późniejsze wybory ludzi, mimo zachowania formalnej możliwości ich zmiany.

Co interesujące, poczucie autonomii nie zawsze zmniejsza wpływ rekomendacji. Eksperymenty nad systemami rekomendacyjnymi wykazały, że zapewnienie użytkownikom większej autonomii wyboru może wręcz zwiększać akceptację sugestii. Jest to istotna korekta intuicji, według której samo zaoferowanie kilku opcji gwarantuje wolność poznawczą. Człowiek może odczuwać autonomię, a jednocześnie coraz częściej wybierać w obrębie zestawu wcześniej skonstruowanego przez system. Autonomia wyboru nie jest tożsama z autonomią konstruowania przestrzeni wyboru. Problem ten jest kluczowy dla generatywnej AI, ponieważ model nie tylko rekomenduje jedną spośród istniejących możliwości, ale może sam wygenerować ich zbiór, wpływając tym samym na autonomią konstruowania przestrzeni wyboru.

Iluzja wyboru a wzrost zależności epistemicznej

Gdy użytkownik pyta o strategię przedsiębiorstwa i otrzymuje pięć wariantów, odnosi realne wrażenie wyboru, choć wszystkie propozycje zostały wygenerowane przez tę samą architekturę, w oparciu o te same dane i ukryte założenia. Pluralizm opcji może zatem maskować monizm generującego je modelu. Człowiek zachowuje wprawdzie swobodę wyboru między wariantami, lecz traci kontakt z rozwiązaniami, których system nie uwzględnił. W epistemologii można to określić jako wzrost zależności epistemicznej (epistemic dependence). Sama w sobie nie jest ona nowa ani patologiczna – nikt nie jest w stanie osobiście zweryfikować całej wiedzy, z której korzysta. Współczesna nauka i społeczeństwo opierają się na podziale pracy poznawczej oraz zaufaniu do ekspertów, instytucji i aparatury badawczej. Artykuł opublikowany w „Synthese” w 2025 roku przekonująco argumentuje, że zależność naukowca od nieprzejrzystego AI nie tworzy całkowicie nowej kategorii epistemologicznej: nauka już wcześniej polegała na narzędziach i specjalistach, których użytkownik nie potrafił w pełni samodzielnie zweryfikować. Nowa może być jednak intensywność i koncentracja tej zależności.

Osoba korzystająca z wiedzy ekspertów ma tradycyjnie możliwość porównywania wielu źródeł, obserwowania sporów i oceny reputacji. System generatywny potrafi natomiast zsyntetyzować wszystkie te warstwy w pojedynczą, płynną odpowiedź, usuwając z pola widzenia konflikt źródeł i samą ścieżkę powstawania wiedzy. W rezultacie użytkownik otrzymuje nie społeczne pole epistemiczne, lecz jego skompresowaną reprezentację. Im bardziej wygodna staje się ta synteza, tym większe ryzyko, że przestanie on rozróżniać wiedzę od jej prezentacji.

Współczesna epistemologia AI wprowadza pojęcie epistemic deference – sytuacji, w której podmiot nie traktuje informacji z AI jedynie jako jednego z dowodów, lecz podporządkowuje jej własny osąd, uznając wyższość systemu. Praca z 2025 roku dotycząca Artificial Epistemic Authorities wskazuje, że taka deferencja może być racjonalna w obliczu trwałej przewagi systemu, staje się jednak problematyczna, gdy prowadzi do „preemption" – czyli zastępowania niezależnych racji użytkownika zamiast ich uzupełniania. Autor proponuje zatem podejście oparte na całkowitym zbiorze dowodów, w którym wynik AI pozostaje istotnym, lecz nie absolutnym składnikiem uzasadnienia. Z praktycznego punktu widzenia oznacza to, że przyszła eksperckość może w mniejszym stopniu polegać na znajomości odpowiedzi, a w większym na zarządzaniu źródłami epistemicznego autorytetu. Ekspert będzie musiał umieć ustalić, które pytania można bezpiecznie delegować, kiedy model wykracza poza swój zakres, jak rozpoznać konflikt między źródłami oraz w których momentach statystyczna trafność AI nie rozstrzyga normatywnego problemu decyzji. Jest to kompetencja metapoznawcza wyższego rzędu. Można jednak łatwo ulec złudzeniu i założyć, że każdy człowiek stanie się w ten sposób „menedżerem AI".

Pozorny nadzór i pułapka odpowiedzialności

Rzeczywistość organizacyjna bywa znacznie bardziej asymetryczna. Jeśli systemy zostaną wdrożone w procesach o wysokim tempie i ogromnej liczbie przypadków, pracownik może nie otrzymać czasu ani informacji niezbędnych do prowadzenia metanadzoru. Jego rola może formalnie obejmować weryfikację, lecz w praktyce sprowadzać się do zwykłego zatwierdzania. Właśnie dlatego kategoria human in the loop jest niewystarczająca. Człowiek może znajdować się w pętli przyczynowej, a jednocześnie nie posiadać meaningful human control.

Materialnie sensowna kontrola wymaga kumulatywnego spełnienia kilku warunków: możliwości uświadomienia sobie, że system może się mylić; dostępu do podstaw decyzji; kompetencji pozwalających zakwestionować rezultat; czasu na przeprowadzenie kontroli; instytucjonalnego uprawnienia do odmowy oraz braku tak silnych bodźców organizacyjnych, które czyniłyby odrzucenie wyniku kosztownym lub praktycznie niemożliwym. Jeśli zabraknie któregokolwiek z

tych elementów, ostatnie kliknięcie człowieka staje się jedynie rytuał legitymizacyjny. Prowadzi to do pojęcia moral crumple zone, wprowadzonego przez Madeleine Clare Elish.

W złożonych systemach automatycznych odpowiedzialność może zostać przypisana najbliższemu człowiekowi nawet wtedy, gdy posiadał on ograniczoną rzeczywistą kontrolę nad przebiegiem zdarzeń. Elish porównuje taką osobę do strefy kontrolowanego zgniotu, która absorbuje siłę zderzenia; w systemie socjotechnicznym człowiek absorbuje jednak odpowiedzialność moralną i prawną za awarię całej architektury. Gujral przejmuje tę kategorię i wiąże ją z problemem projektowania systemów hybrydowych, wskazując, że nie można przypisywać pełnej odpowiedzialności osobie, której realna zdolność kontrolowania automatycznego systemu została ograniczona.

Zależność między kontrolą a odpowiedzialnością jest centralna dla sprawczości. Sprawczość nie polega wyłącznie na byciu fizycznym źródłem ostatniego działania; wymaga uczestnictwa w konstruowaniu celu, rozumienia sytuacji i zdolności modyfikowania przebiegu zdarzeń. Dlatego człowiek zatwierdzający rekomendację AI może być formalnym decydentem, a zarazem posiadać bardzo słabą sprawczość epistemiczną. Odwrotnie – model nieposiadający podmiotowości moralnej może wykazywać ogromną sprawczość przyczynową poprzez kształtowanie informacji i możliwości dostępnych ludziom. Powstaje asymetria: sprawczość przyczynowa ulega rozproszeniu, podczas gdy odpowiedzialność prawna pozostaje skoncentrowana. Najnowsze eksperymenty nad przypisywaniem odpowiedzialności w zespołach człowiek-AI potwierdzają, że ocena winy zależy nie tylko od obiektywnej roli systemu, ale również od wcześniejszej wiedzy uczestników oraz postrzeganej wagi zadania.

Jest to kolejny argument za odrzuceniem prostej formuły: „człowiek zawsze odpowiada, bo podjął końcową decyzję”. Odpowiedzialność wymaga rekonstrukcji architektury informacyjnej i decyzyjnej, a nie tylko wskazania ostatniego punktu interakcji. Jeszcze poważniejszą konsekwencję może mieć pętla uczenia człowiek-AI. W tradycyjnym modelu wyobrażamy sobie AI jako zewnętrzne narzędzie o określonych biasach, a człowieka jako niezależny podmiot, który może je skorygować. W rzeczywistości użytkownik uczy się podczas kontaktu z systemem. Badania Glickmana i Sharot obejmujące 1401 uczestników wykazały, że powtarzana interakcja z obciążonym biasem AI może zmieniać ludzkie osądy percepcyjne, emocjonalne i społeczne, prowadząc do amplifikacji uprzedzeń po stronie człowieka.

Nie mamy więc do czynienia z dwoma statycznymi komponentami, lecz z układ współadaptacyjny. Zasadniczo zmienia to teorię bezpieczeństwa. Jeśli model wpływa na człowieka, a zachowania tego ostatniego stają się później danymi dla kolejnych modeli lub systemów rekomendacyjnych, może powstać sprzężenie zwrotne utrwalające określone reprezentacje rzeczywistości. Gujral nazywa

takie błędy duchami w pętli i ostrzega, że powtarzane przez AI sposoby wnioskowania mogą stawać się dla użytkownika nowymi standardami myślenia. Nauka pozwala tę metaforę uściślić: ryzykiem nie jest tajemnicze przenikanie błędu z maszyny do podświadomości, lecz uczenie społeczne, ekspozycja, kotwiczenie, normalizacja oraz wzajemna adaptacja systemu rekomendacyjnego i użytkownika. Błędem jest zatem traktowanie biasu AI i biasu człowieka jako niezależnych składników, które jedynie się sumują. Ich interakcja ma charakter nieliniowy.

AI może wzmacniać ludzkie uprzedzenia, człowiek dostarczać systemowi danych potwierdzających wcześniejszą rekomendację, a organizacja interpretować rosnącą zgodność jako dowód jakości modelu. Powstaje wtedy epistemiczna pętla pozytywnego sprzężenia, w której zmniejszanie różnicy między człowiekiem a maszyną jest błędnie interpretowane jako wzrost trafności, podczas gdy w rzeczywistości oba komponenty konwergują ku temu samemu błędowi.

Ryzyko homogenizacji poznawczej i pułapka pozornej wyjaśnialności

Pojawia się tutaj głębszy problem kulturowy. System generatywny może stopniowo wpływać nie tylko na pojedyncze decyzje, ale na sam styl uzasadniania. Jeśli miliony osób korzystają z podobnych modeli przy tworzeniu raportów, analiz, komunikatów, prac naukowych czy dokumentów prawnych, system zaczyna pośredniczyć w reprodukcji form argumentacyjnych. Nie musi narzucać jednej ideologii, aby homogenizować repertuar retoryczny, długość wywodu, sposób formułowania kontrargumentów czy zestaw najłatwiej dostępnych analogii. Zależność kognitywna może zatem zyskać wymiar kulturowy szybciej niż polityczny. Właśnie dlatego epistemiczny pluralizm powinien zostać uznany za element bezpieczeństwa. Osoba podejmująca decyzję przy wsparciu jednego systemu powinna w sytuacjach wysokiej stawki mieć możliwość uzyskania niezależnego modelu problemu, konkurencyjnych źródeł lub krytycznego wglądu drugiego systemu. Nie chodzi o generowanie sztucznego sporu dla samego pluralizmu, lecz o zapobieganie sytuacji, w której jeden model definiuje jednocześnie fakty, alternatywy, kryteria oceny i rekomendację. Architektura, w której wszystkie te poziomy są kontrolowane przez ten sam system, może być niezwykle spójna, lecz spójność nie jest synonimem prawdy.

Należy zachować ostrożność również wobec explainable AI. Intuicyjnie wydaje się, że im więcej wyjaśnień, tym większa sprawczość człowieka, jednak nie zawsze jest to prawdą. Wyjaśnienie może zwiększać zaufanie do systemu nawet wtedy, gdy jest niepełne, trudne do zweryfikowania lub jedynie przekonujące retorycznie. Przegląd automation bias z 2025 roku wskazuje, że wpływ wyjaśnialności na nadmierne poleganie na AI zależy od jakości, złożoności i sposobu prezentacji

informacji; sama transparentność nie stanowi uniwersalnego antidotum. W systemach generatywnych problem ten się pogłębia, ponieważ model może stworzyć płynne uzasadnienie, które nie jest wierną reprezentacją mechanizmu prowadzącego do odpowiedzi. Dlatego w systemie hybrydowym należy odróżnić wyjaśnienie perswazyjne od wyjaśnienia epistemicznego. Pierwsze zwiększa jedynie poczucie zrozumienia, drugie zaś dostarcza informacji realnie pozwalających ocenić niezawodność wyniku. W praktyce użytkownik często potrzebuje mniej narracji, a więcej danych o niepewności, źródłach, warunkach brzegowych i kontrprzykładach. System mówiący długo i przekonująco może ograniczać sprawczość epistemiczną bardziej niż system podający zwięzłą odpowiedź z wyraźnym zaznaczeniem: „tej części nie wiem". W tym kontekście epistemic agency można zdefiniować jako zdolność podmiotu do aktywnego uczestniczenia w tworzeniu, ocenie i rewizji własnych przekonań.

Autonomia to umiejętność zarządzania zależnością

Autonomia ta nie wymaga samowystarczalności. Człowiek może polegać na ekspertach, narzędziach i AI, zachowując sprawczość, o ile rozumie strukturę zależności i posiada możliwość krytycznej rewizji. Traci ją natomiast wtedy, gdy jego przekonania są systematycznie kształtowane przez źródło, którego rekomendacji nie potrafi ocenić, nie może go obejść i nie rozpoznaje nawet jako zewnętrznego wpływu. Epistemiczna autonomia nie jest więc brakiem zależności; jest zdolnością zarządzania zależnością. Stąd wynika inna koncepcja przyszłego human-centered AI.

Zadaniem sprawnego systemu nie powinno być maksymalizowanie liczby przypadków, w których człowiek akceptuje jego odpowiedź. Taka metryka nagradzałaby system za perswazyjność. Celem powinno być podniesienie jakości decyzji człowieka przy jednoczesnym zachowaniu jego zdolności do krytycznej kontroli. W niektórych sytuacjach oznacza to podanie rekomendacji, w innych przedstawienie kilku alternatyw, a w jeszcze innych celowe powstrzymanie się od odpowiedzi i wskazanie, że decyzja wymaga rodzaju wiedzy lub wartościowania, którego system nie posiada. W tym właśnie miejscu idea proaktywnego tarcia z książki Gujrala może zostać przekształcona z metafory w zasadę projektową. Tarcie powinno być proporcjonalne do ryzyka, odwracalności i epistemicznej asymetrii.

Nie ma sensu zmuszać użytkownika do głębokiej refleksji przy każdej korekcie pisowni. Inaczej należy projektować system przygotowujący decyzję medyczną, prawną, finansową albo dotyczącą użycia siły. Im większa stawka i mniej odwracalny skutek, tym istotniejsze stają się alternatywy, niezależne źródła, jawne wskazywanie niepewności i możliwość zatrzymania procesu. Właściwie zaprojektowane tarcie nie jest więc antytezą efektywności. Jest alokacją wysiłku poznawczego tam,

gdzie marginalna wartość refleksji jest największa. Człowiek nie powinien wykonywać wszystkiego samodzielnie; powinien zachować wysiłek dla tych punktów procesu, w których decyzja jest szczególnie wrażliwa na błąd, wartości albo nietypowy kontekst. AI mogłaby wręcz pomagać określać, kiedy takie punkty występują, pod warunkiem że sama procedura eskalacji nie jest całkowicie zależna od modelu, którego błąd ma wykrywać.

Jest to zapowiedź nowej teorii podziału pracy poznawczej. W epoce przemysłowej podstawowym problemem była alokacja czynności fizycznych i organizacyjnych. W epoce AI kluczowa staje się alokacja funkcji epistemicznych: kto obserwuje, kto pamięta, kto formułuje hipotezę, kto szacuje prawdopodobieństwo, kto ustanawia wartość, kto rozstrzyga konflikt i kto odpowiada za konsekwencję. Nie jest przy tym konieczne, aby wszystkie funkcje pozostawały po stronie człowieka. Konieczne jest natomiast, aby architektura ich podziału była jawna i nie obarczała człowieka odpowiedzialnością za etap procesu, którego realnie nie kontroluje. W tej perspektywie największym zagrożeniem nie jest "AI, która zaczyna myśleć", lecz człowiek, który stopniowo przestaje rozpoznawać, które elementy jego własnego myślenia powstały poza nim. Granica pomiędzy sugestią a przekonaniem, przypomnieniem a pamięcią, rekomendacją a preferencją może stać się coraz bardziej płynna. Im lepiej system poznaje użytkownika, tym skuteczniej dostosowuje formę informacji do jego wcześniejszych reakcji. Personalizacja, która zwiększa użyteczność, równocześnie potęguje zdolność systemu do wpływania na warunki powstawania sądu.

Nie oznacza to konieczności obrony jakiegoś historycznie czystego, "niezapośredniczonego" umysłu. Taki umysł nigdy nie istniał. Człowiek zawsze myślał językiem odziedziczonym po innych, za pomocą pojęć kultury, w strukturach instytucjonalnych i przy użyciu narzędzi. Nowość polega na interaktywności, adaptacyjności i skali pośrednika. Książka nie zmienia swojej argumentacji w reakcji na profil psychologiczny czytelnika; model może to robić. Biblioteka nie przestawia półek po każdej decyzji badacza; system rekomendacyjny może modyfikować przyszłe środowisko informacji właśnie na podstawie wcześniejszych wyborów.

Dlatego hybrydowa kognicja powinna być rozumiana jako dynamiczny układ dwóch zmieniających się systemów, a nie połączenie stałego człowieka ze stałym narzędziem. Człowiek uczy się AI, AI adaptuje się do człowieka, organizacja przekształca procedury wokół AI, a następnie nowe procedury modyfikują sposób kształcenia kolejnego pokolenia ludzi. W kilku cyklach technologia może zmienić nie tylko wykonanie zawodu, ale również kryterium tego, co w danym zawodzie uchodzi za kompetencję. To jest właśnie koewolucja kognitywna. Jeżeli systemy te mają zwiększać, a nie rozpuszczać ludzką sprawczość, należy odejść od modelu "human in the loop" na rzecz modelu human agency in the loop. Pierwszy sprawdza obecność człowieka; drugi pyta, czy obecność ta ma rzeczywiste znaczenie przyczynowe i epistemiczne. Czy człowiek posiada niezależną reprezentację

problemu? Czy może wykryć błąd? Czy odmowa jest możliwa? Czy decyzja zostawia ślad pozwalający później ustalić, skąd pochodziły jej przesłanki? Czy użytkownik zachowuje kompetencję po latach pracy z systemem? Dopiero te pytania pozwalają ocenić, czy system jest rozszerzeniem ludzkiego osądu, czy jego substytutem ukrytym pod fasadą nadzoru. W tym miejscu dochodzimy do punktu, w którym sama analiza delegacji przestaje wystarczać. Skoro człowiek i AI wzajemnie modyfikują swoje procesy, trzeba opisać pozytywną architekturę ich połączenia. Nie każdy outsourcing poznawczy prowadzi do deskillingu; nie każda integracja prowadzi do utraty autonomii.

Możliwe są konfiguracje, w których człowiek i system wzmacniają swoje odmienne przewagi, zachowując możliwość wzajemnej korekty.

Delegowanie osądu jako zagrożenie dla sprawczości

Właśnie temu służą koncepcje Centaura, Cyborga, Fused Mind, Hive, extended mind, distributed cognition oraz human-centered AI. Pozwalają one przejść od diagnozy zależności do teorii hybrydowego systemu poznawczego: określić, kiedy człowiek i AI tworzą rzeczywistą nową jednostkę funkcjonalną, jakie typy sprzężenia są możliwe i gdzie przebiega granica między rozszerzeniem podmiotowości a jej stopniowym przejęciem przez infrastrukturę. To przejście od delegowania zadań do delegowania osądu, obejmujące zjawiska cognitive offloading, automation bias oraz przebudowę ludzkiej sprawczości.

Najważniejszym problemem w relacji człowieka ze sztuczną inteligencją nie jest sama automatyzacja pracy, lecz przesuwanie granicy między tym, co wykonujemy osobiście, a tym, co pozostaje integralnym składnikiem naszego procesu poznawczego.

Można bowiem powierzyć maszynie rachunek, pozostając autorem problemu, kryterium rozwiązania i końcowej oceny wyniku. Można powierzyć jej wyszukanie informacji, zachowując samodzielność interpretacji. Można jednak przekazać również określanie tego, które informacje są istotne, generowanie hipotez, prognozowanie ryzyka, hierarchizację alternatyw czy rekomendację działania, a następnie przygotowanie uzasadnienia, które pozwoli przedstawić decyzję jako racjonalną. W tym ostatnim przypadku człowiek formalnie nadal może być osobą "podejmującą decyzję", lecz coraz większa część struktury poznawczej poprzedzającej wybór znajduje się poza nim. Właśnie tutaj przebiega granica pomiędzy automatyzacją czynności a delegowaniem osądu.

Psychologia poznawcza od dawna zna zjawisko cognitive offloading, czyli przenoszenia części operacji poznawczych na zewnętrzne artefakty lub środowisko. Człowiek zapisuje termin spotkania,

by nie obciążać pamięci; korzysta z mapy, aby nie rekonstruować trasy; wykonuje rachunek na kalkulatorze zamiast przechowywać wyniki pośrednie w pamięci roboczej. Takie odciążanie nie jest aberracją, lecz podstawową cechą ludzkiej inteligencji. Rozszerzone poznanie rozwija się właśnie dlatego, że ograniczony biologiczny system potrafi modyfikować środowisko tak, aby zmniejszyć koszt własnych operacji.

Problem pojawia się dopiero wtedy, gdy narzędzie przejmuje nie tylko obciążenie pamięciowe lub obliczeniowe, lecz również funkcje niezbędne do rozpoznania, czy wynik zasługuje na akceptację. Generatywna AI zasadniczo zmienia skalę tego zjawiska, ponieważ zewnętrzny system nie ogranicza się już do przechowywania dostarczonej przez człowieka treści. Może tworzyć własną syntezę, ustalać strukturę dokumentu, wybierać argumenty, sugerować priorytety, przygotowywać kontrargumenty, wykrywać domniemane ryzyka, redagować komunikat i rekomendować następne działanie.

System zewnętrzny przesuwa się zatem w górę hierarchii poznawczej: od pamięci pomocniczej ku zewnętrznemu generatorowi reprezentacji problemu. To przesunięcie jest istotniejsze niż prosty wzrost produktywności, ponieważ decyzja człowieka zależy od tego, jak wcześniej została skonstruowana przestrzeń alternatyw. Proces ten można ująć poprzez metaforę "duchów w pętli sprzężenia zwrotnego" oraz ostrzeżenie przed nadmiernie gładką współpracą z AI. Gujral zwraca uwagę, że system może wpływać na ludzkie poznanie poprzez subtelną modulację procesów poprzedzających refleksyjny osąd, a brak "tarcia" poznawczego może prowadzić do sytuacji, w której człowiek zaczyna raczej konsumować gotową myśl niż samodzielnie angażować się w jej wytwarzanie. W ujęciu analitycznym należy jednak odrzucić zbyt prostą wizję, według której AI stopniowo "wyłącza System 2". Rzeczywisty mechanizm jest bardziej złożony: technologia zmienia koszt poszczególnych strategii poznawczych, a człowiek – zgodnie z zasadami ograniczonej racjonalności – adaptuje się do nowej struktury tych kosztów.

Deskilling jako koszt krótkoterminowej efektywności

Jeśli samodzielne przeszukiwanie źródeł zajmuje godzinę, a otrzymanie syntezy trwa kilka sekund, korzystanie z tej drugiej opcji jest racjonalne. Podobnie, gdy przygotowanie pierwszej wersji argumentacji wymaga znacznego wysiłku, a AI dostarcza ją natychmiast, jej wykorzystanie wydaje się uzasadnione.

Problem nie zaczyna się zatem od "lenistwa poznawczego". Pojawia się on w momencie, gdy krótkoterminowo racjonalna strategia oszczędzania energii zmienia długofalową strukturę kompetencji. Jednostka może systematycznie wybierać optymalną lokalnie ścieżkę współpracy z AI,

by po kilku latach odkryć, że posiada mniejszą zdolność samodzielnego wykonywania operacji niezbędnych do nadzoru nad tą samą technologią. To klasyczny konflikt między doraźną efektywnością a kapitałem poznawczym.

Zjawisko to najlepiej opisuje pojęcie deskilling. Nie oznacza ono ogólnego spadku inteligencji, lecz utratę lub osłabienie konkretnej umiejętności wskutek rzadszego jej samodzielnego stosowania. Przegląd dotyczący wykorzystania AI w medycynie z 2026 roku wskazał już empiryczne przesłanki takiego ryzyka: w części badań zaobserwowano pogorszenie jakości samodzielnej pracy po okresie korzystania z systemów AI, podatność diagnostów na błędne sugestie oraz spadek liczby przypadków, na których młodzi specjaliści mogliby ćwiczyć niezależną diagnozę. Autorzy podkreślali, że choć dowody są wciąż ograniczone, pozostają one na tyle spójne, by uznać ochronę kompetencji za element bezpieczeństwa klinicznego.

Taka sytuacja odsłania szczególnie trudny paradoks kompetencji. Dzisiejszy ekspert zazwyczaj uczył się zawodu w świecie, w którym AI odgrywała marginalną rolę; posiada więc samodzielny model dziedziny i traktuje technologię jako rozszerzenie nabytej wiedzy. Następne pokolenie profesjonalistów może natomiast zdobywać kompetencje już w środowisku, w którym część zadań jest od początku zautomatyzowana. Badanie jakościowe z 2025 roku ukazało właśnie ten rozdźwięk: doświadczeni pracownicy ochrony zdrowia nie dostrzegali u siebie silnego deskillingu, lecz wyrażali obawę, że młodsi klinicyści mogą nie mieć wystarczająco wielu okazji do wypracowania niezależnego osądu, skoro część pracy diagnostycznej przejmie AI.

Nie jest to zjawisko ograniczone wyłącznie do medycyny.

Augmentacja zamiast automatyzacji chroni kompetencje

W programowaniu generowanie kodu może ograniczać liczbę sytuacji, w których początkujący musi samodzielnie przejść od problemu do struktury rozwiązania. W prawie automatyczne streszczenia orzecznictwa mogą skracać proces wypracowywania umiejętności czytania i porównywania źródeł. W administracji model tworzący projekt decyzji może redukować liczbę przypadków, w których urzędnik samodzielnie rekonstruuje podstawę prawną. Z kolei w nauce automatyczna synteza literatury może ograniczać ekspozycję na sprzeczności i słabe sygnały obecne w źródłach pierwotnych. Efekt zależy od sposobu użycia: ta sama technologia może przyspieszać rozwój eksperckości, jeśli pełni rolę nauczyciela i partnera krytycznego, albo ją osłabiać, jeśli konsekwentnie eliminuje poznawczo formacyjne etapy pracy. Dlatego opozycja augmentation

versus automation jest ważniejsza niż proste pytanie o to, czy AI została "użyta". Augmentacja zwiększa możliwości człowieka, zachowując lub rozwijając kompetencje niezbędne do sprawowania kontroli nad wynikiem. Automatyzacja natomiast może optymalizować sam rezultat, niezależnie od rozwoju operatora. W krótkim okresie oba modele mogą produkować identycznie dobrą decyzję; różnica między nimi ujawnia się jednak w czasie, gdy system przestaje działać, pojawia się sytuacja nietypowa lub algorytm popełnia błąd.

Organizacja oparta na augmentacji zachowuje rezerwę ludzkich kompetencji. Organizacja oparta na głębokiej substytucji może odkryć, że formalnie obecny człowiek nie potrafi już samodzielnie odtworzyć procesu, który ma nadzorować. Na tym tle należy analizować automation bias – tendencję do nadmiernego polegania na rekomendacjach systemów zautomatyzowanych, co objawia się przyjmowaniem błędnych sugestii lub pomijaniem informacji, których system nie wskazał. Najnowszy przegląd 35 badań dotyczących human-AI collaboration pokazuje, że automation bias nie wynika z jednego mechanizmu. Wpływają na niego m.in. poziom wiedzy użytkownika, AI literacy, charakter zadania, złożoność wyjaśnienia, dynamika zaufania, koszt weryfikacji oraz indywidualne różnice poznawcze. Właściwym problemem nie jest zatem samo "zaufanie do AI", lecz kalibracja zaufania względem rzeczywistej wiarygodności systemu w konkretnej klasie sytuacji. Błędem jest również zbyt małe zaufanie: jeśli system jest niezawodnie lepszy od człowieka w określonej domenie, uporczywe odrzucanie jego rekomendacji obniża jakość decyzji. Istnieje zatem zarówno automation bias, jak i algorithm aversion.

Celem dobrze zaprojektowanej współpracy nie jest maksymalizowanie ludzkiej niezależności, lecz uzyskanie racjonalnej dystrybucji epistemicznego autorytetu. Człowiek powinien wiedzieć, w jakich sytuacjach system jest statystycznie lepszy, gdzie zawodzi, jakie informacje wykorzystuje i jaką wagę nadać jego rekomendacji. Jest to problem zaufania warunkowego, a nie ogólnego. Najbardziej pouczające są przypadki, w których człowiek posiada prawidłowy sąd początkowy, a następnie zmienia go pod wpływem błędnej sugestii AI. W badaniach klinicznych od dawna obserwuje się takie odwrócenia; nowsze dane potwierdzają, że błędna rekomendacja może istotnie pogarszać trafność decyzji. Badanie opublikowane w 2026 roku wykazało, że poprawne rekomendacje AI podnosiły wyniki diagnostyczne, podczas gdy błędne wyraźnie je obniżały – kluczowym czynnikiem ochronnym pozostawała własna kompetencja profesjonalisty. Oznacza to, że system hybrydowy nie może być oceniany poprzez niezależną dokładność człowieka i modelu. Interakcja tworzy trzeci układ poznawczy, którego właściwości nie są prostą sumą komponentów. Możemy to zapisać jako problem interakcyjnej epistemologii. Jeżeli człowiek ma rację w 80 procentach przypadków, a AI w 90 procentach, nie wynika z tego automatycznie, że zespół człowiek-AI osiągnie wynik powyżej 90 procent.

Może go osiągnąć, jeśli człowiek potrafi rozpoznać momenty, w których model się myli, i vice versa. Może jednak osiągnąć wynik gorszy od obu elementów, jeśli rekomendacje systemu destabilizują trafne sądy człowieka lub człowiek ingeruje w prawidłowe wyniki modelu. Jakość współpracy zależy więc przede wszystkim od komplementarności błędów oraz mechanizmu arbitrażu.

Proaktywne tarcie jako narzędzie ochrony niezależności poznawczej

W tym miejscu pojawia się pojęcie cognitive forcing, czyli celowego projektowania interfejsu tak, aby spowolnić automatyczne przyjęcie rekomendacji i wymusić pewien stopień niezależnego przetwarzania. Można na przykład poprosić użytkownika o wstępną ocenę przed pokazaniem sugestii AI, opóźnić wyświetlenie wyniku lub wymagać wskazania argumentu przemawiającego przeciwko propozycji systemu. Gujral proponuje podobny mechanizm pod nazwą "proaktywnego tarcia", wiążąc go z przedstawianiem alternatyw, ujawnianiem niepewności, możliwością wycofania działania i obowiązkiem autoryzacji decyzji o wysokiej stawce. Jednocześnie najnowsze badania wskazują, że samo spowolnienie procesu nie jest panaceum.

W eksperymentach dotyczących oceny ryzyka przemocy w sytuacjach kryzysu psychiatrycznego stronnicze rekomendacje AI prowadziły do bardziej stronniczych decyzji uczestników, a zastosowane techniki cognitive forcing nie wyeliminowały tego efektu. Większą odporność wykazywały osoby o wysokiej skłonności do angażowania się w wysiłkowe przetwarzanie poznawcze, choć mechanizm ten zależy również od zmęczenia i kontekstu. Wniosek jest teoretycznie istotny: tarcie poznawcze działa tylko wtedy, gdy prowadzi do realnego przetworzenia alternatywy, a nie staje się kolejnym formalnym kliknięciem. Nie wystarczy zatem dodać przycisku "czy jesteś pewwna?". Użytkownik może szybko nauczyć się klikać go automatycznie, podobnie jak akceptuje regulaminy czy ostrzeżenia bezpieczeństwa. Projektowanie autonomii wymaga stworzenia sytuacji, w której informacja niezbędna do krytycznej oceny rzeczywiście staje się poznawczo dostępna. Czasami lepszym rozwiązaniem będzie pokazanie argumentów przeciw rekomendacji, przedstawienie kilku konkurujących hipotez, ujawnienie, że przypadek wykracza poza typowy rozkład treningowy, a niekiedy całkowite powstrzymanie się AI od udzielenia jednej odpowiedzi. Kluczowa jest tutaj kolejność informacji.

Jeśli człowiek najpierw otrzymuje stanowisko AI, jego późniejsza analiza nie rozpoczyna się w stanie neutralnym. Rekomendacja staje się kotwicą, wpływa na interpretację dowodów i ustanawia domyślną hipotezę. Dlatego rozwiązanie, w którym profesjonalista dokonuje niezależnej oceny przed ujawnieniem wyniku modelu, może chronić epistemiczną różnorodność zespołu. Nie jest to

zawsze praktyczne ani efektywne, ale pokazuje istotną zależność: jeżeli chcemy uzyskać komplementarność dwóch systemów poznawczych, nie powinniśmy doprowadzać do sytuacji, w której drugi system natychmiast podporządkowuje pierwszy własnemu punktowi wyjścia. Problem ustawień domyślnych rozciąga się daleko poza decyzje eksperckie. Psychologia wyboru od dziesięcioleci wskazuje na znaczenie default effects: opcja wstępnie wybrana posiada przewagę wynikającą z bezwładu, kosztu zmiany, interpretacji rekomendacji i ograniczonej uwagi. W systemach AI ustawieniem domyślnym może być nie tylko zaznaczone pole, ale pierwsza wygenerowana odpowiedź. Jeśli system prezentuje jedną propozycję jako naturalny rezultat analizy, alternatywy zyskują status czegoś wymagającego dodatkowego wysiłku. Badania z 2026 roku dotyczące algorytmicznych ustawień domyślnych w decyzjach społecznych wykazały, że ekspozycja na określone predefiniowane preferencje algorytmu może wpływać na późniejsze wybory ludzi, mimo zachowania formalnej możliwości ich zmiany.

Iluzja wyboru a ryzyko uległości epistemicznej

Gdy użytkownik pyta o strategię przedsiębiorstwa i otrzymuje pięć wariantów, odnosi realne poczucie wyboru, choć wszystkie propozycje pochodzą z tej samej architektury, opierają się na tych samych danych i ukrytych założeniach. Pluralizm opcji może zatem maskować monizm generującego je modelu. Człowiek zachowuje wprawdzie swobodę wyboru między wariantami, lecz traci kontakt z rozwiązaniami, których system nie wygenerował. W epistemologii można to określić jako wzrost zależności epistemicznej (epistemic dependence). Sama w sobie nie jest ona nowa ani patologiczna – nikt nie jest w stanie osobiście zweryfikować całej wiedzy, z której korzysta. Współczesna nauka i społeczeństwo opierają się na podziale pracy poznawczej oraz zaufaniu do ekspertów, laboratoriów, instytucji i aparatury. Artykuł opublikowany w Synthese w 2025 roku przekonująco argumentuje, że zależność naukowca od nieprzejrzystego AI nie tworzy całkowicie nowej kategorii epistemologicznej: nauka już wcześniej polegała na narzędziach i autorytetach, których użytkownik nie potrafił w pełni samodzielnie zweryfikować. Nowa może być jednak intensywność i koncentracja tej zależności.

Osoba korzystająca z wiedzy ekspertów ma tradycyjnie możliwość porównywania wielu źródeł, śledzenia sporów i oceny reputacji. System generatywny potrafi natomiast syntetyzować wszystkie te warstwy w pojedynczą, płynną odpowiedź, usuwając z pola widzenia konflikt źródeł i samą ścieżkę powstawania wiedzy. W rezultacie użytkownik otrzymuje nie społeczne pole epistemiczne, lecz jego skompresowaną reprezentację. Im wygodniejsza staje się ta synteza, tym większe ryzyko, że przestanie on rozróżniać wiedzę od jej prezentacji.

Współczesna epistemologia AI wprowadza pojęcie epistemic deference – sytuacji, w której podmiot nie tylko traktuje informację z AI jako jeden z dowodów, ale podporządkowuje jej własny osąd, uznając wyższość systemu. Praca z 2025 roku dotycząca Artificial Epistemic Authorities wskazuje, że taka deferencja może być racjonalna wobec systemu wykazującego trwałą przewagę, lecz staje się problematyczna, gdy prowadzi do „preemption" – zastępowania niezależnych racji użytkownika zamiast ich uzupełniania. Autor proponuje zatem podejście oparte na całkowitym zbiorze dowodów, w którym wynik AI pozostaje istotnym, lecz nie absolutnym składnikiem uzasadnienia. Z praktycznego punktu widzenia oznacza to, że przyszła eksperckość może w mniejszym stopniu polegać na znaniu odpowiedzi, a w większym na zarządzaniu źródłami epistemicznego autorytetu. Ekspert będzie musiał umieć ustalić, które pytania można bezpiecznie delegować, kiedy model wykracza poza swój zakres, jak rozpoznać konflikt między źródłami oraz w których momentach statystyczna trafność AI nie rozstrzyga normatywnego problemu decyzji. Jest to kompetencja metapoznawcza wyższego rzędu. Można jednak łatwo przecenić tę wizję i założyć, że każdy człowiek stanie się „menedżerem AI".

Iluzja kontroli i ryzyko epistemicznej pętli sprzężenia

Rzeczywistość organizacyjna bywa bardziej asymetryczna. Jeśli systemy zostaną wdrożone w procesach o wysokim tempie i ogromnej liczbie przypadków, pracownik może nie otrzymać czasu ani informacji niezbędnych do prowadzenia metanadzoru. Jego rola formalnie obejmie weryfikację, lecz w praktyce sprowadzi się do zwykłego zatwierdzania. Właśnie dlatego kategoria human in the loop jest niewystarczająca. Człowiek może znajdować się w pętli przyczynowej, a jednocześnie nie posiadać meaningful human control.

Materialnie sensowna kontrola wymaga kumulatywnego spełnienia kilku warunków: świadomości, że system może się mylić; dostępu do podstaw decyzji; kompetencji pozwalających zakwestionować rezultat; czasu na przeprowadzenie kontroli; instytucjonalnego uprawnienia do odmowy oraz braku tak silnych bodźców organizacyjnych, które czyniłyby odrzucenie wyniku kosztownym lub praktycznie niemożliwym. Jeśli zabraknie choć jednego z tych elementów, ostatnie kliknięcie człowieka może stanowić jedynie rytuał legitymizacyjny. Prowadzi to do pojęcia moral crumple zone, wprowadzonego przez Madeleine Clare Elish.

W złożonych systemach automatycznych odpowiedzialność bywa przypisana najbliższemu człowiekowi nawet wtedy, gdy jego rzeczywista kontrola nad przebiegiem zdarzeń była ograniczona. Elish porównuje taką osobę do strefy kontrolowanego zgniotu, która absorbuje siłę zderzenia; w systemie socjotechnicznym człowiek absorbuje jednak odpowiedzialność moralną i prawną za

awarię całej architektury. Gujral wiąże tę kategorię z problemem projektowania systemów hybrydowych, wskazując, że nie można obarczać pełną odpowiedzialnością kogoś, czyja realna zdolność kontrolowania automatyzacji została ograniczona.

Zależność między kontrolą a odpowiedzialnością jest centralna dla sprawczości. Sprawczość nie polega wyłącznie na byciu fizycznym źródłem ostatniego działania; wymaga uczestnictwa w konstruowaniu celu, rozumienia sytuacji i zdolności modyfikowania przebiegu zdarzeń. Dlatego osoba zatwierdzająca rekomendację AI może być formalnym decydentem, a zarazem posiadać bardzo słabą sprawczość epistemiczną. Odwrotnie – model pozbawiony podmiotowości moralnej może mieć ogromną sprawczość przyczynową, kształtując informacje i możliwości dostępne ludziom. Powstaje asymetria: sprawczość przyczynowa ulega rozproszeniu, podczas gdy odpowiedzialność prawna pozostaje skoncentrowana. Najnowsze eksperymenty nad przypisywaniem odpowiedzialności w zespołach człowiek-AI potwierdzają, że ocena winy zależy nie tylko od obiektywnej roli systemu, ale również od wcześniejszej wiedzy uczestników oraz postrzeganej wagi zadania.

Jest to kolejny argument za odrzuceniem prostej formuły: „człowiek zawsze odpowiada, bo podjął końcową decyzję". Odpowiedzialność wymaga rekonstrukcji architektury informacyjnej i decyzyjnej, a nie tylko wskazania ostatniego punktu interakcji. Jeszcze poważniejszą konsekwencję może mieć pętla uczenia człowiek-AI. W tradycyjnym modelu AI postrzegamy jako zewnętrzne narzędzie z określonymi biasami, a człowieka jako niezależny podmiot zdolny do ich korekty. W rzeczywistości użytkownik uczy się podczas kontaktu z systemem. Badania Glickmana i Sharot na grupie 1401 uczestników wykazały, że powtarzana interakcja z obciążonym biasem AI może zmieniać ludzkie osądy percepcyjne, emocjonalne i społeczne, prowadząc do amplifikacji uprzedzeń po stronie człowieka.

Nie mamy więc do czynienia z dwoma statycznymi komponentami, lecz z układem współadaptacyjnym. Zmienia to zasadniczo teorię bezpieczeństwa. Jeśli model wpływa na człowieka, a zachowania tego ostatniego stają się później danymi dla kolejnych modeli lub systemów rekomendacyjnych, powstaje sprzężenie zwrotne utrwalające określone reprezentacje rzeczywistości. Gujral nazywa takie błędy „duchami w pętli" i ostrzega, że powtarzane przez AI sposoby wnioskowania mogą stawać się dla użytkownika nowymi standardami myślenia. Nauka pozwala tę metaforę uściślić: ryzykiem nie jest tajemnicze przenikanie błędu z maszyny do podświadomości, lecz uczenie społeczne, ekspozycja, kotwiczenie, normalizacja oraz wzajemna adaptacja systemu i użytkownika. Niebezpiecznym błędem jest zatem traktowanie biasu AI i biasu człowieka jako niezależnych składników, które jedynie się sumują; ich interakcja ma charakter nieliniowy.

Pluralizm epistemiczny przeciwko homogenizacji myślenia

Pojawia się tutaj głębszy problem kulturowy. System generatywny może stopniowo wpływać nie tylko na pojedyncze decyzje, ale na sam styl uzasadniania. Jeśli miliony osób korzystają z podobnych modeli przy tworzeniu raportów, analiz, komunikatów czy prac naukowych i prawnych, system zaczyna pośredniczyć w reprodukcji konkretnych form argumentacyjnych. Nie musi on narzucać jednej ideologii, aby homogenizować repertuar retoryczny, długość wywodu, sposób formułowania kontrargumentów czy zestaw najłatwiej dostępnych analogii. Zależność kognitywna może zatem zyskać wymiar kulturowy jeszcze przed wymiarem politycznym.

Właśnie dlatego pluralizm epistemiczny powinien zostać uznany za element bezpieczeństwa. Osoba podejmująca decyzję przy wsparciu jednego systemu powinna – zwłaszcza w sytuacjach o wysokiej stawce – mieć możliwość odwołania się do niezależnego modelu problemu, konkurencyjnych źródeł lub krytycznej roli drugiego systemu. Nie chodzi tu o generowanie sztucznego sporu dla samego pluralizmu, lecz o zapobieganie sytuacji, w której jeden model definiuje jednocześnie fakty, alternatywy, kryteria oceny i finalną rekomendację. Architektura, w której wszystkie te poziomy są kontrolowane przez ten sam system, może być niezwykle spójna, jednak spójność nie jest synonimem prawdy.

Należy zachować ostrożność również wobec koncepcji explainable AI (XAI). Intuicyjnie wydaje się, że im szersze wyjaśnienie, tym większa sprawczość człowieka, lecz nie zawsze jest to prawdą. Wyjaśnienie może bowiem zwiększać zaufanie do systemu nawet wtedy, gdy jest niepełne, trudne do zweryfikowania lub jedynie przekonujące retorycznie. Przegląd automation bias z 2025 roku wskazuje, że wpływ wyjaśnialności na nadmierne poleganie na AI zależy od jakości, złożoności i sposobu prezentacji informacji; sama transparentność nie stanowi uniwersalnego antidotum. W systemach generatywnych problem ten się pogłębia, gdyż model potrafi wygenerować płynne uzasadnienie, które nie jest wierną reprezentacją mechanizmu prowadzącego do odpowiedzi.

Dlatego w systemie hybrydowym należy odróżnić wyjaśnienie perswazyjne od wyjaśnienia epistemicznego. Pierwsze zwiększa jedynie poczucie zrozumienia; drugie dostarcza informacji, które realnie pozwalają ocenić niezawodność wyniku. W praktyce użytkownik często potrzebuje mniej narracji, a więcej danych o niepewności, źródłach, warunkach brzegowych i kontrprzykładach. System mówiący długo i przekonująco może ograniczać sprawczość epistemiczną bardziej niż system podający zwięzłą odpowiedź z wyraźnym zastrzeżeniem: „tej części nie wiem”. W tym kontekście epistemic agency można zdefiniować jako zdolność podmiotu do aktywnego uczestniczenia w tworzeniu, ocenie i rewizji własnych przekonań.

Od formalnego nadzoru do realnej sprawczości epistemicznej

Autonomia nie wymaga samowystarczalności. Człowiek może polegać na ekspertach, narzędziach i AI, zachowując sprawczość, o ile rozumie strukturę tych zależności i posiada możliwość krytycznej rewizji. Traci ją natomiast wtedy, gdy jego przekonania są systematycznie kształtowane przez źródło, którego rekomendacji nie potrafi ocenić, nie może go obejść, a co gorsza – nie rozpoznaje nawet jako zewnętrznego wpływu. Epistemiczna autonomia nie jest więc brakiem zależności, lecz zdolnością do zarządzania tą zależnością. Stąd wynika nowa koncepcja human-centered AI.

Zadaniem sprawnego systemu nie powinno być maksymalizowanie liczby przypadków, w których człowiek akceptuje podaną odpowiedź; taka metryka nagradzałaby system za perswazyjność. Celem winno być podniesienie jakości decyzji człowieka przy jednoczesnym zachowaniu jego zdolności do krytycznej kontroli. W zależności od kontekstu oznacza to podanie konkretnej rekomendacji, przedstawienie kilku alternatyw lub celowe powstrzymanie się od odpowiedzi i wskazanie, że decyzja wymaga wiedzy bądź wartościowania, którego system nie posiada. To właśnie tutaj idea proaktywnego tarcia może zostać przekształcona z metafory w zasadę projektową. Tarcie powinno być proporcjonalne do ryzyka, odwracalności skutków i epistemicznej asymetrii.

Nie ma sensu wymuszać długiej refleksji przy każdej korekcie pisowni. Inaczej należy projektować systemy wspierające decyzje medyczne, prawne, finansowe czy dotyczące użycia siły. Im wyższa stawka i mniej odwracalny skutek, tym kluczniejsze stają się alternatywy, niezależne źródła, jawne wskazywanie niepewności oraz możliwość zatrzymania procesu. Właściwie zaprojektowane tarcie nie jest zatem antytezą efektywności; jest alokacją wysiłku poznawczego tam, gdzie marginalna wartość refleksji jest największa. Człowiek nie powinien wykonywać wszystkiego samodzielnie, lecz zachować wysiłek dla tych punktów procesu, w których decyzja jest szczególnie wrażliwa na błąd, wartości lub nietypowy kontekst. AI mogłaby wręcz pomagać identyfikować takie momenty, pod warunkiem że sama procedura eskalacji nie zależy całkowicie od modelu, którego błąd ma wykrywać.

Jest to zapowiedź nowej teorii podziału pracy poznawczej. W epoce przemysłowej kluczowa była alokacja czynności fizycznych i organizacyjnych; w epoce AI najważniejsza staje się alokacja funkcji epistemicznych: kto obserwuje, kto pamięta, kto formułuje hipotezę, szacuje prawdopodobieństwo, ustanawia wartość, rozstrzyga konflikt i odpowiada za konsekwencje. Nie jest konieczne, aby wszystkie te funkcje pozostały po stronie człowieka. Konieczne jest jednak, by architektura ich podziału była jawna i nie obarczała człowieka odpowiedzialnością za etap procesu, którego realnie nie kontroluje. W tej perspektywie największym zagrożeniem nie jest „AI, która zaczyna myśleć”,

lecz człowiek, który stopniowo przestaje rozpoznawać, które elementy jego własnego myślenia powstały poza nim. Granica między sugestią a przekonaniem, przypomnieniem a pamięcią, rekomendacją a preferencją staje się płynna. Im lepiej system zna użytkownika, tym skuteczniej dostosowuje formę informacji do jego reakcji. Personalizacja zwiększająca użyteczność jednocześnie wzmacnia zdolność systemu do wpływania na warunki powstawania sądu.

Nie oznacza to jednak konieczności obrony mitycznego, „niezapośredniczonego" umysłu – taki nigdy nie istniał. Człowiek zawsze myślał językiem odziedziczonym, za pomocą pojęć kultury i w strukturach instytucjonalnych. Nowością jest interaktywność, adaptacyjność i skala pośrednika. Książka nie zmienia argumentacji w reakcji na profil psychologiczny czytelnika; model AI potrafi to robić. Biblioteka nie przestawia półek po każdej decyzji badacza; system rekomendacyjny może modyfikować całe środowisko informacyjne na podstawie wcześniejszych wyborów.

Dlatego hybrydowa kognicja powinna być rozumiana jako dynamiczny układ dwóch zmieniających się systemów, a nie połączenie stałego człowieka ze stałym narzędziem. Człowiek uczy się AI, AI adaptuje się do człowieka, organizacja przekształca procedury wokół technologii, a nowe procedury kształtują sposób edukacji kolejnych pokoleń. W kilku cyklach technologia może zmienić nie tylko wykonanie zawodu, ale i kryterium tego, co w nim uchodzi za kompetencję. To jest właśnie koewolucja kognitywna. Jeśli systemy te mają zwiększać, a nie rozpuszczać ludzką sprawczość, należy odejść od modelu „human in the loop" na rzecz human agency in the loop. Pierwszy sprawdza jedynie obecność człowieka; drugi pyta, czy obecność ta ma rzeczywiste znaczenie przyczynowe i epistemiczne. Czy człowiek posiada niezależną reprezentację problemu? Czy może wykryć błąd? Czy odmowa jest możliwa? Czy decyzja zostawia ślad pozwalający ustalić pochodzenie jej przesłanek? Czy użytkownik zachowuje kompetencje po latach pracy z systemem? Dopiero te pytania pozwalają ocenić, czy system jest rozszerzeniem ludzkiego osądu, czy jego substytutem ukrytym pod fasadą nadzoru. W tym miejscu analiza delegacji przestaje wystarczać. Skoro człowiek i AI wzajemnie modyfikują swoje procesy, konieczne jest opisanie pozytywnej architektury ich połączenia. Nie każdy outsourcing poznawczy prowadzi do deskillingu; nie każda integracja kończy się utratą autonomii.

Ku teorii hybrydowego systemu poznawczego

Możliwe są konfiguracje, w których człowiek i system wzmacniają swoje odmienne przewagi, zachowując przy tym zdolność do wzajemnej korekty. Właśnie temu służą koncepcje Centaura, Cyborga, Fused Mind, Hive, extended mind, distributed cognition oraz human-centered AI. Kolejna część musi zatem przejść od diagnozy zależności ku teorii hybrydowego systemu

poznawczego: określić, w jakich warunkach człowiek i AI tworzą rzeczywistą nową jednostkę funkcjonalną, jakie typy sprzężenia są możliwe i gdzie przebiega granica między rozszerzeniem podmiotowości a jej stopniowym przejęciem przez infrastrukturę.

> W świecie coraz gładszej współpracy z algorytmami największym ryzykiem nie jest bunt maszyn, lecz powolne znikanie człowieka z własnego procesu myślowego. Czy w pogoni za efektywnością nie budujemy systemu, w którym będziemy jedynie świadkami decyzji, których nie potrafimy już ani zakwestionować, ani w pełni zrozumieć? Ostatecznie pytanie o autonomię sprowadza się do tego, czy wciąż wiemy, gdzie kończy się sugestia modelu, a zaczyna nasze własne przekonanie.

Inspiracje

[1] **"The AI Instinct"** Rana Gujral

Rana Gujral (ur. 18 czerwca 1976 r.) to amerykański przedsiębiorca technologiczny, dyrektor wykonawczy i inwestor, specjalizujący się w kognitywnej sztucznej inteligencji (AI) i sztucznej inteligencji emocjonalnej (AI). Najbardziej znany jest jako dyrektor generalny Behavioral Signals, firmy zajmującej się sztuczną inteligencją, która opracowuje technologie rozpoznawania emocji i analizy mowy, aby wyodrębniać intencje, nastroje i sygnały behawioralne z wzorców głosowych. Przed objęciem stanowiska dyrektora Behavioral Signals, Gujral założył firmę TiZE, zajmującą się oprogramowaniem do uczenia maszynowego, i kierował nią aż do jej przejęcia przez Alchemy. Zajmował również kluczowe stanowiska kierownicze, przyczyniając się do restrukturyzacji firm z branży elektroniki użytkowej, takich jak Cricut i Logitech. Jako wybitny mówca i komentator branżowy zajmujący się odpowiedzialnym rozwojem AI, kognitywistyką człowiek-maszyna oraz sztuczną inteligencją ogólną, Gujral publikował liczne artykuły w najważniejszych publikacjach biznesowych i na globalnych forach technologicznych.

Rana Gujral *(born June 18, 1976) is an American technology entrepreneur, executive, and investor specializing in cognitive artificial intelligence and emotion AI. He is best known as the Chief Executive Officer of Behavioral Signals, an enterprise artificial intelligence company that develops emotion-recognition and speech-analytics technology to extract intent, sentiment, and behavioral cues from vocal patterns. Prior to leading Behavioral Signals, Gujral founded the machine learning software company TiZE, guiding it until its acquisition by Alchemy, and held key leadership roles driving corporate turnaround at consumer electronics companies including Cricut and Logitech. A prominent speaker and industry commentator on responsible AI development, human-machine cognition, and artificial general intelligence, Gujral has contributed widely to major business publications and global technology forums.*

Behavioral Signals

Literatura uzupełniająca

Artykuły naukowe

Prace autorów przywoływanych w tekście:

[1] "Real-World Integration of a Sepsis Deep Learning Technology Into Routine Clinical Care: Implementation Study"

Mark Sendak, William Ratliff, Dina Sarro, Elizabeth Alderton, Joseph Futoma, Madeleine Clare Elish

2019 | JMIR Medical Informatics | DOI: 10.2196/15182

Źródło

[2] "“This Seems to Work”: Designing Technological Systems with The Algorithmic Imaginations of Those Who Labor"

Lindsey D. Cameron, Angèle Christin, Michael Ann DeVito, Tawanna R. Dillahunt, M. C. Elish

2021 | CHI Extended Abstracts | DOI: 10.1145/3411763.3441331

Źródło

[3] "How Choice Reveals and Shapes Expected Hedonic Outcome"

Tali Sharot, Benedetto De Martino, Raymond J. Dolan

2009 | Journal of Neuroscience | DOI: 10.1523/jneurosci.4972-08.2009

Źródło

[4] "How human–AI feedback loops alter human perceptual, emotional and social judgements"

Moshe Glickman, Tali Sharot

2024 | Nature Human Behaviour | DOI: 10.1038/s41562-024-02077-2

Źródło

[5] "AI-assisted discovery of unknown cognitive biases: Humans systematically undervalue information under complexity"

Moshe Glickman, Tali Sharot

2026 | DOI: 10.31234/osf.io/vabzc_v1

Źródło

Artykuły naukowe autora:

[6] "The Nature of Human Cognition and AI's Approach to Mimicking It"

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch02

Źródło

[7] **"AI in the Real World: Applications and Use Cases"**
RANA GUJRAL, KAI-FU LEE
2026 | DOI: 10.1002/9781394392155.ch03
Źródło

[8] **"The AI Instinct"**
RANA GUJRAL, KAI-FU LEE
2026 | DOI: 10.1002/9781394392155.fmatter
Źródło

[9] **"The Emergence of Artificial General Intelligence (AGI)"**
RANA GUJRAL, KAI-FU LEE
2026 | DOI: 10.1002/9781394392155.ch04
Źródło

[10] **"Introduction: Our Definition of “There” Is Wrong"**
RANA GUJRAL, KAI-FU LEE
2026 | DOI: 10.1002/9781394392155.ch01
Źródło

Treść tworzy Fundacja Dobre Państwo.
Redaguje i publikuje APA ONE, autorski system redakcyjny Fundacji oparty na AI.
APA ONE Publishing System
Fundacja Dobre Państwo © 2026
Article ID: f07b3b8e-72ba-4511-a918-e62a9d874db4