Architektury kognicji sztucznej i ich implementacje systemowe w świetle koncepcji The AI Instinct

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Tekst analizuje naturę sztucznej inteligencji nie jako kopię ludzkiego umysłu, lecz jako zbiór różnorodnych architektur poznawczych. Autor kontrastuje podejście symboliczne, oparte na jawnych regułach i manipulacji reprezentacjami, z koneksjonizmem, który stawia na rozproszone reprezentacje wiedzy wyłaniające się w procesie uczenia statystycznego. Omówiono ewolucję od systemów ekspertowych, przez głębokie sieci neuronowe, aż po transformery i mechanizm uwagi. Analiza wskazuje na przejście od sztywnych ontologii do elastycznych modeli generatywnych, podkreślając różnice w sposobie generalizacji, pamiętania i planowania informacji w zależności od przyjętej architektury systemowej.

The text analyzes the nature of artificial intelligence not as a copy of the human mind, but as a set of diverse cognitive architectures. The author contrasts the symbolic approach, based on explicit rules and representation manipulation, with connectionism, which relies on distributed knowledge representations emerging through statistical learning. It discusses the evolution from expert systems, through deep neural networks, to transformers and the attention mechanism. The analysis points to a transition from rigid ontologies to flexible generative models, emphasizing differences in how information is generalized, remembered, and planned depending on the adopted system architecture.

Artykuł rzuca wyzwanie powszechnemu przekonaniu, że sztuczna inteligencja jest jedynie dążącą do perfekcji kopią ludzkiego umysłu. Autor stawia tezę, że AI stanowi odrębną, modularną architekturę poznawczą, która nie replikuje biologicznej psychologii, lecz buduje własny reżim przetwarzania informacji. Kluczowym punktem

analizy jest przejście od postrzegania AI jako narzędzia do uznania jej za krytyczną infrastrukturę kognitywną. Systemy te stopniowo przejmują fundamentalne funkcje percepcji, prognozowania i działania w sektorach takich jak medycyna, finanse czy obronność, co prowadzi do powstania głębokiej zależności instytucjonalnej. W tym kontekście koncepcja Artificial General Experience (AGE) okazuje się istotniejsza niż samo AGI; to nie poziom abstrakcyjnych kompetencji, lecz ciągłość interakcji z rzeczywistością i budowanie stabilnej historii operacyjnej definiują nową formę maszynowego sprawstwa. Tekst dowodzi, że zamiast pytać o świadomość AI, powinniśmy badać moment, w którym systemy te stają się niezbędnym elementem społecznego łańcucha decyzyjnego, trwale przekształcając sposób, w jaki ludzkość postrzega i przetwarza świat.

SŁOWA KLUCZOWE

architektury kognicji sztucznej | implementacje systemowe | The AI Instinct | modele świata | agentowość techniczna | systemy multimodalne | pamięć parametryczna | wnioskowanie przyczynowe | infrastruktura poznawcza | large reasoning models | ucieleśniony agent | systemic dependency | moral crumple zone | kognicja rozproszona

AI jako odrębna rodzina architektur poznawczych

Natura kognicji sztucznej: od manipulacji symbolami do modeli generatywnych, agentów i sztucznych modeli świata

Porównanie ludzkiej kognicji ze sztuczną inteligencją wymaga odrzucenia dwóch symetrycznych uproszczeń. Pierwsze z nich traktuje AI jako „niedokończonej kopii człowieka", której kolejne generacje mają stopniowo uzupełniać braki biologicznego umysłu. Drugie zakłada, że mamy do czynienia z tym samym rodzajem systemu poznawczego co u ludzi, tyle że osadzonym w innym substracie i dysponującym większą szybkością przetwarzania. Oba te podejścia zbyt wcześnie przesądzają o wyniku analizy.

Sztuczna inteligencja nie wyewoluowała z organizmu dążącego do zachowania własnej integralności, lecz powstała jako zbiór projektowanych metod rozwiązywania problemów. Nie posiada jednej genealogii, która odpowiadałaby rozwojowi biologicznego układu nerwowego. Jej

historia to splot logiki formalnej, statystyki, teorii sterowania i informacji, optymalizacji, programowania symbolicznego, sieci neuronowych czy współczesnych systemów generatywnych. Z tego powodu właściwsze jest mówienie o „rodzinie sztucznych architektur poznawczych" niż o jednej, spójnej naturze AI. Dopiero analiza tego, jak poszczególne architektury reprezentują informacje, generalizują, pamiętają i planują, pozwala ocenić, w jakim stopniu pojęcia zapożyczone z psychologii człowieka zachowują w tym kontekście sens.

Najwcześniejsze ambitne programy AI wyrastały z intelektualnego gruntu rewolucji poznawczej i komputacjonizmu. Choć konferencja w Dartmouth z 1956 roku stała się symbolicznym początkiem instytucjonalnej AI, jej program nie powstał w próżni. Poprzedzały go teoria automatów, cybernetyka Norberta Wienera, formalizacja obliczalności oraz rozwój pierwszych komputerów cyfrowych. W paradygmacie symbolicznym inteligentne zachowanie miało wynikać z operowania reprezentacjami według jawnych reguł: program otrzymywał symbole reprezentujące elementy problemu i przekształcał je zgodnie z procedurą, by dojść do formalnie opisywalnego rozwiązania. Podejście to sprawdzało się znakomicie w zadaniach o jednoznacznie zdefiniowanych granicach – przy dowodzeniu twierdzeń, rozwiązywaniu łamigłówek czy w systemach ekspertowych opartych na regułach dziedzinowych.

Znaczenie symbolicznej AI dla filozofii poznania przewyższyło jej praktyczne sukcesy. Udowodniła ona, że zachowania utożsamiane z inteligencją można sprowadzić do manipulacji formalnymi reprezentacjami. Jednocześnie obnażyła fundamentalną trudność: świat rzeczywisty nie jest dostarczany systemowi w formie gotowych, jednoznacznych symboli. Rozpoznanie istotnych informacji, znaczenia kontekstu czy selekcja kluczowych konsekwencji działania okazały się znacznie trudniejsze niż samo stosowanie reguł po uprzednim zakodowaniu problemu. Trudność ta przejawiała się w różnych formach – od problemu ramy po brak odporności systemów ekspertowych na sytuacje wykraczające poza ręcznie zdefiniowaną wiedzę. System mógł być doskonałym specjalistą w świecie nakreślonym przez programistę, lecz stawał się „niezwykle kruchym agentem", gdy rzeczywistość rozmijała się z ontologią wpisaną w jego reguły.

Konekcjonizm i wyłanianie się rozproszonych reprezentacji wiedzy

Konekcjonizm zaproponował inną drogę. Zamiast zapisywać struktury wiedzy w postaci jawnych symboli i reguł, system można uczyć na przykładach, modyfikując parametry rozproszonej sieci tak, aby określone wejścia prowadziły do pożądanych wyników. Sieci neuronowe, choć inspirowane ogólnymi zasadami biologicznego przetwarzania, nie są cyfrowymi kopiami mózgu. Ich jednostki,

funkcje aktywacji czy algorytmy optymalizacji to konstrukcje matematyczne podporządkowane celowi obliczeniowemu. Historycznie spór między podejściem symbolicznym a koneksjonistycznym dotyczył czegoś więcej niż techniki: pytał o to, czy inteligencja powinna wynikać z jawnego reprezentowania zasad, czy z wyłaniania się użytecznych reprezentacji w procesie uczenia. Rozwój głębokiego uczenia przesunął ciężar praktycznej AI zdecydowanie w stronę tego drugiego rozwiązania.

Sieci wielowarstwowe nauczyły się wydobywać hierarchie cech bez konieczności ręcznego opisywania każdej z nich przez projektanta. Sukcesy w rozpoznawaniu obrazów, mowy czy tłumaczeniu pokazały, że kompetencje, które długo uznawano za domenę rozbudowanej wiedzy symbolicznej, mogą powstać poprzez uczenie statystycznych regularności w ogromnych zbiorach danych. Nie oznaczało to jednak końca symboli i reguł, lecz zmianę miejsca, w którym lokuje się wiedza systemu. W klasycznym programie zależności można było odczytać bezpośrednio z kodu lub bazy reguł; w głębokiej sieci kompetencje zostają rozproszone pomiędzy parametry i aktywacje, co czyni interpretację wewnętrznych mechanizmów znacznie trudniejszą. Transformery, wprowadzone w 2017 roku, otworzyły kolejną fazę tej ewolucji.

Mechanizm uwagi umożliwił efektywne modelowanie relacji między elementami sekwencji, a skalowanie architektur i mocy obliczeniowej doprowadziło do powstania wielkich modeli językowych. Ich podstawowa zasada może wydawać się zwodniczo prosta: przewidywanie kolejnych elementów sekwencji na podstawie kontekstu. Jednak zadanie to, realizowane na wystarczająco złożonych danych, wymusza uczenie licznych regularności języka, świata przedstawionego w tekstach, relacji semantycznych czy konwencji społecznych. Nie oznacza to, że system posiada wewnętrzną encyklopedię złożoną z jawnych zdań. Wiedza jest zakodowana w rozproszonej strukturze parametrów, a odpowiedź stanowi wynik dynamicznego przetwarzania bieżącego kontekstu. Dlatego stwierdzenie, że model językowy „tylko przewiduje następne słowo", jest prawdziwe na poziomie mechanizmu generatywnego, lecz mylące w kontekście zakresu reprezentacji, jakie system musi wytworzyć, by to zadanie realizować. Podobny błąd popełnilibyśmy, twierdząc, że mózg „tylko generuje potencjały czynnościowe" i uznając rozmowę o pamięci czy planowaniu za zbędną.

Wyjaśnienie mechanizmu na poziomie operacji elementarnych nie zastępuje wyjaśnienia organizacji funkcjonalnej wyższego rzędu. Jednocześnie nie wolno ulec przeciwstawnemu błędowi: złożone zachowanie modelu nie dowodzi automatycznie posiadania kategorii psychicznych właściwych człowiekowi. W tym punkcie należy skorygować jedną z centralnych metafor.

Notatka wielokrotnie przedstawia współczesną AI jako próbę replikacji ludzkiego Systemu 2 – analitycznej i logicznej strony poznania – przeciwstawiając jej biologiczny System 1, obejmujący

intuicję i afektywną selekcję. Intuicja tego rozróżnienia pozostaje cenna: sztuczne systemy nie mają biologicznej historii, która u ludzi wytworzyła szybkie reakcje homeostatyczne. Nie można jednak opisać współczesnych LLM po prostu jako „czystego Systemu 2”. Przegląd z Nature Reviews Psychology z 2025 roku wskazuje, że modele mogą przejawiać zarówno zachowania podobne do odpowiedzi typu System 1 – szybkie heurystyki i systematyczne biasy – jak i przejawy kontrolowanej deliberacji, szczególnie przy zastosowaniu odpowiednich metod inferencji. Autorzy podkreślają jednak, że analogia z ludzkimi systemami dual-process jest ograniczona, gdyż źródła błędów maszynowych mogą być odmienne od mechanizmów biologicznych. Zamiast pytać, czy AI posiada System 1 lub 2, trafniej jest badać reżimy przetwarzania. Model może wygenerować natychmiastową odpowiedź opartą na silnym wzorcu, ale może również wykorzystać dodatkowy budżet inferencyjny, rozłożyć problem na podproblemy i przeprowadzić kolejne iteracje.

Różnica funkcjonalna przypomina rozróżnienie między odpowiedzią automatyczną a deliberatywną, lecz nie wynika z biologicznego przełączania między dwiema rodzinami procesów poznawczych.

Rozdźwięk między funkcjonalnym rozumowaniem a ucieleśnieniem

Wynik jest produktem architektury modelu, treningu, sposobu wywołania, dostępnego czasu obliczeniowego oraz otaczającego systemu wykonawczego. W 2026 roku w literaturze pojawił się już termin "large reasoning models" w odniesieniu do systemów optymalizowanych pod kątem planowania, deliberacji i wieloetapowego rozwiązywania problemów, jednak nawet w nich sprawne rozumowanie współistnieje z charakterystycznymi klasami błędów. Pojęcie "reasoning" wymaga zatem precyzji. W badaniach nad AI może ono oznaczać zdolność do wykonywania zadań opartych na dedukcji, matematyce, planowaniu czy wykorzystywaniu informacji pośrednich i wieloetapowych zależności. Nie rozstrzyga to jednak, czy system "rozumuje" w tym samym sensie fenomenologicznym lub psychologicznym co człowiek. Co więcej, wygenerowane uzasadnienie nie musi być wiernym protokołem mechanizmu, który doprowadził do wyniku. Model może tworzyć tekst przedstawiający logiczną argumentację tylko dlatego, że taka argumentacja jest odpowiednim rezultatem generacyjnym, a nie dlatego, że otrzymujemy dzięki niej transparentny zapis jego wewnętrznego przetwarzania. Rozróżnienie między zdolnością rozwiązania problemu, umiejętnością wygenerowania przekonującego uzasadnienia a faktyczną interpretowalnością mechanizmu staje się zatem jednym z kluczowych wymogów epistemologii AI.

Dalsze komplikacje wprowadza multimodalność. Klasyczny model językowy operuje na reprezentacjach tekstowych; system multimodalny potrafi natomiast integrować tekst z obrazem, dźwiękiem, wideo i innymi sygnałami. Nie oznacza to automatycznego powstania maszynowego odpowiednika ludzkiej percepcji, lecz zasadniczo poszerza zakres struktur, które model może reprezentować. Badania nad wizualnymi zdolnościami multimodalnych LLM wykazują zarówno imponujące podobieństwa funkcjonalne do pewnych ludzkich kompetencji poznawczych, jak i systematyczne różnice dowodzące, że maszyny stosują inne strategie rozwiązywania zadań niż ludzie. Rozwój dużych modeli multimodalnych zmierza ku coraz bardziej zintegrowanym mechanizmom uczenia i generowania ponad granicami tekstu, obrazu i wideo, co zaciera dawny podział między odrębnymi systemami językowymi a percepcyjnymi. Multimodalność nie jest jednak tożsama z ucieleśnieniem. Kamera dostarczająca obraz nie czyni jeszcze systemu organizmem posiadającym perspektywę cielesną.

Ucieleśniony agent musi znajdować się w pętli działania: jego ruch zmienia środowisko, środowisko generuje nowe dane, a te z kolei wpływają na dalsze działanie. Z tego punktu widzenia szczególnie istotne są robotyka, vision-language-action models oraz systemy sterujące urządzeniami fizycznymi, gdyż przesuwają one AI z domeny biernego przetwarzania zapisanych danych w stronę interakcji z otwartym światem. To właśnie architekturom ucieleśnionym przypisuje się szczególne znaczenie jako potencjalnej drodze ku AGE, ponieważ fizyczna interakcja tworzy ciągłość między percepcją, działaniem a konsekwencją. Tezę tę należy jednak rozumieć funkcjonalnie: robot może uczyć się konsekwencji swoich działań bez dowodu na to, że posiada fenomenalne doświadczenie tychże konsekwencji.

Modele świata i pamięć jako fundamenty agentowości

Tu pojawia się pojęcie modelu świata. W najszerszym znaczeniu jest to wewnętrzna struktura pozwalająca systemowi przewidywać, jak stan środowiska zmieni się pod wpływem działania lub upływu czasu. Modele świata są kluczowe dla planowania – agent nie musi bowiem testować każdej potencjalnej czynności w rzeczywistości, lecz może symulować jej możliwe następstwa. W uczeniu ze wzmocnieniem od dawna rozróżnia się podejścia model-free, uczące polityk działania lub wartości bez jawnego modelowania dynamiki środowiska, od model-based, gdzie przewidywanie stanów wspomaga wybór. Współczesna debata o world models rozszerza tę intuicję na systemy generatywne zdolne do modelowania złożonych struktur przestrzennych, czasowych i przyczynowych.

Nie należy jednak utożsamiać modelu świata z „rozumieniem świata”. System może skutecznie przewidywać trajektorie bez posiadania pojęciowego modelu przyczynowości, podobnie jak potrafi odwzorowywać statystyczne regularności języka bez jawnego reprezentowania praw przyrody. Judea Pearl zwracał uwagę na różnicę między obserwacją, interwencją a rozumowaniem kontrfaktycznym; sama predykcja korelacyjna nie jest jeszcze pełnym wnioskowaniem przyczynowym.

To rozróżnienie jest kluczowe dla AI operującej w świecie. Agent może trafnie przewidywać, co zwykle następuje po określonym zdarzeniu, lecz w sytuacji wymagającej odpowiedzi na pytanie „co stanie się, jeśli celowo zmienię X?”, potrzebuje silniejszej reprezentacji struktury przyczynowej. Jeszcze wyższy poziom stanowi zdolność kontrfaktyczna – rozważenie tego, co mogłoby się wydarzyć, gdyby wcześniejszy przebieg zdarzeń był inny. Gujral sytuuje modele świata i wnioskowanie przyczynowe jako jedne z najważniejszych dróg prowadzących od samej kompetencji ku bardziej ciągłemu systemowi działania. Jego mocniejsza teza, wedle której system posiadający modele świata zbliża się do „doświadczenia”, wymaga jednak dalszej analizy. Predykcja następstwa, reprezentacja przyczynowa, pamięć interakcji, aktualizacja zachowania na podstawie szkody oraz fenomenalne przeżycie konsekwencji to odmienne właściwości.

AGE będzie wartościową teorią tylko wtedy, gdy nie zostaną one zamknięte w jednym, nieostrym pojęciu. Analogicznego porządku wymaga kwestia pamięci AI. W ludzkiej kognicji rozróżniamy wiele systemów: m.in. pamięć roboczą, epizodyczną, semantyczną i proceduralną. W AI słowo „pamięć” obejmuje równie zróżnicowane mechanizmy, które często błędnie traktuje się jak funkcjonalne odpowiedniki jednego ludzkiego procesu. Pierwszym poziomem jest pamięć parametryczna: regularności zakodowane w wagach modelu podczas treningu. Drugim – bieżący kontekst inferencyjny, pozwalający wykorzystywać informacje podane w danej interakcji. Trzecim są zewnętrzne systemy retrieval, bazy wiedzy i dokumenty pobierane doraźnie. Czwartym może być trwała pamięć użytkownika lub agenta, zachowywana między sesjami. Piątym – historia działania zapisywana i wykorzystywana do późniejszej modyfikacji strategii. Formy te różnią się mechanizmem, trwałością i stopniem, w jakim zdarzenie z przeszłości zmienia przyszłe zachowanie.

Taki podział pozwala precyzyjniej sformułować intuicję materiału, wedle której doświadczenie powinno oznaczać coś więcej niż storage. System może dysponować ogromną bazą wspomnień i pozostawać strukturalnie niezmieniony przez ich treść. Może również ulec trwałej aktualizacji po jednym zdaniu bez zachowania narracyjnego wspomnienia tego epizodu. Biologiczna pamięć łączy te procesy w sposób nieporównywalnie bardziej spleciony; w systemie sztucznym można je projektować oddzielnie.

Dlatego pytanie o AGE nie powinno brzmieć „czy AI ma pamięć?”, lecz: jakiego rodzaju przeszłość jest dostępna systemowi, jak wpływa ona na jego przyszłe decyzje i czy tworzy ciągłość pozwalającą mówić o historycznie rozwijającym się agencie? Właśnie z połączenia modelu językowego z pamięcią, narzędziami i pętlą działania wyrasta współczesna agentowość AI. Sam model generujący odpowiedź pozostaje w dużej mierze reaktywny: otrzymuje wejście i tworzy rezultat. Agent natomiast może otrzymać cel, rozłożyć go na czynności, korzystać z wyszukiwarki lub kodu, wykonywać operacje w zewnętrznym środowisku, obserwować skutki i planować dalsze kroki. W literaturze z 2025 i 2026 roku przejście od pojedynczego LLM do systemów agentowych opisuje się coraz częściej poprzez trzy funkcje – rozumowanie, działanie i interakcję – oraz integrację planowania, pamięci, narzędzi i informacji zwrotnej.

Agentowość techniczna to autonomia działania, a nie sprawczość osobowa

Nie mamy tu do czynienia z prostym zwiększeniem mocy modelu, lecz ze zmianą samej jednostki technicznej, której zachowanie analizujemy. Agentowość techniczna nie powinna być jednak bezrefleksyjnie utożsamiana ze sprawczością osobową. Agent programowy może realizować długie sekwencje działań i dostosowywać je do środowiska, nie posiadając przy tym własnych celów w sensie psychologicznym. Jego autonomiczność oznacza przede wszystkim rzadsze interwencje człowieka między początkiem a końcem procesu. Jest to cecha niezwykle istotna z punktu widzenia praktyki, lecz ontologicznie skromniejsza niż autonomia moralna. DeepMind w swoim frameworku „Levels of AGI” celowo rozdziela szerokość i poziom kompetencji od problemu autonomii wdrożeniowej, wskazując, że ogólność zdolności oraz stopień samodzielności działania to dwa odrębne wymiary. System może być zatem wysoce kompetentny, lecz silnie nadzorowany, lub mniej kompetentny, a jednocześnie posiadać szeroki zakres samodzielnych działań w środowisku.

Wraz z wprowadzeniem agentów zmienia się również charakter błędu. W tradycyjnym modelu generatywnym pomyłka pozostaje jedynie tekstem na ekranie. W systemie agentowym ta sama błędna reprezentacja może stać się przesłanką do kolejnych kroków: modyfikować pliki, uruchamiać procesy, przekazywać informacje innym agentom i wpływać na realne zasoby. Ryzyko nie jest więc prostą funkcją prawdopodobieństwa halucynacji; zależy ono od długości horyzontu działania, dostępu do narzędzi, odwracalności decyzji oraz mechanizmów kontroli pozwalających wykryć błąd, zanim ulegnie on kumulacji.

W systemach wieloagentowych pojawia się dodatkowo problem emergencji zachowań wynikających z interakcji kilku modeli. Literatura z 2026 roku zwraca uwagę, że zespoły agentów potrafią

wykonywać złożone zadania dzięki komunikacji i podziałowi ról, co jednak zwiększa wymóg transparentności całego procesu. W tym miejscu model Roju otrzymuje bardziej konkretną interpretację. Centaur oznacza podział ról między człowiekiem a AI, Cyborg silniejszą integrację, natomiast Hive to zbiorowa inteligencja, w której rośnie ryzyko outsourcingu osądu. Z perspektywy współczesnej architektury agentowej typologię tę można rozwinąć: Centaur zachowuje stosunkowo czytelne granice kompetencji, system cyborgiczny opiera się na częstych sprzężeniach zwrotnych między człowiekiem a modelem, Rój zaś przenosi część inteligencji na strukturę koordynacji wielu komponentów.

W takim układzie pojedynczy agent nie musi rozumieć całego procesu, podobnie jak pracownik w złożonej organizacji nie musi znać pełnej ścieżki powstawania decyzji. Inteligencja staje się własnością organizacji przepływów. Pojawia się tu analogia do kognicji rozproszonej Hutchinsa, jednak nie należy jej stosować mechanicznie. W załodze okrętu role pełnią ludzie posiadający własną odpowiedzialność, wiedzę i społeczne rozumienie sytuacji. W systemie wieloagentowym role są jedynie wygenerowanymi instrukcjami, a poszczególni agenci mogą dzielić te same źródła błędów, dane treningowe i ograniczenia. Wielość agentów nie gwarantuje zatem epistemicznego pluralizmu – dziesięć modeli z tej samej rodziny może zgodnie powielać tę samą błędną reprezentację. Liczba uczestników nie jest więc równoznaczna z różnorodnością poznawczą społeczeństwa.

Funkcjonalna autoreprezentacja to nie samoświadomość

Podobny problem dotyczy funkcji refleksji i „samokrytyki” AI. System można skonstruować tak, aby wygenerował pierwsze rozwiązanie, następnie je skrytykował, poszukał kontrprzykładów i poprawił rezultat. Funkcjonalnie przypomina to kontrolę metapoznawczą, jednak nie oznacza to, że model posiada introspekcyjny dostęp analogiczny do ludzkiego. W wielu przypadkach druga generacja jest po prostu kolejnym procesem inferencyjnym, w którym pierwszy wynik służy jako wejście. Może to prowadzić do znaczącej poprawy jakości, ale również do racjonalizacji błędnej odpowiedzi. Dlatego pojęcia reflection, self-critique czy self-evaluation w literaturze AI należy odczytywać jako nazwy procedur, a nie bezpośrednie dowody istnienia samoświadomości.

Należy tutaj oddzielić model siebie od jaźni. Agent może potrzebować reprezentacji własnych możliwości: wiedzieć, z jakich narzędzi potrafi korzystać, jakie posiada zasoby, czego mu brakuje, ile czasu pozostało do wykonania zadania i jakie działania podjął wcześniej. Taki self-model może

znacząco zwiększać skuteczność systemu, lecz nie wynika z niego automatycznie fenomenalne poczucie bycia sobą.

Z punktu widzenia teorii AGE rozróżnienie to jest zasadnicze, ponieważ Gujral umieszcza „model jaźni" i odpowiednik Default Mode Network w projektowanym Experience Stack. Wersja naukowo ostrożniejsza powinna mówić o warstwie samomodelowania i monitorowania własnych stanów – dopóki brakuje podstaw, by przejść od funkcjonalnej autoreprezentacji do twierdzeń o podmiotowym doświadczeniu. Nieprecyzyjne byłoby również stwierdzenie, że modele językowe nie mają żadnego „kontaktu z rzeczywistością". Współczesne systemy mogą korzystać z wyszukiwarek, baz danych, sensorów, wykonania kodu, urządzeń fizycznych i informacji zwrotnej ze środowiska. Problem tkwi raczej w charakterze tego kontaktu i sposobie aktualizacji wewnętrznej reprezentacji. Model bez narzędzi generuje twierdzenia na podstawie regularności z treningu i kontekstu; model wyposażony w retrieval może sprawdzić źródło; agent może wykonać eksperyment obliczeniowy, a robot fizycznie oddziaływać na otoczenie.

Są to jakościowo różne pętle epistemiczne. Mówienie ogólnie o „AI" zaciera różnice większe niż te, które w psychologii dzielą odmienne mechanizmy poznawcze. Z tego powodu ocena sztucznej kognicji wymaga taksonomii funkcji zamiast pojedynczej skali inteligencji. Możemy badać zdolność do transferu między domenami, uczenia nowych zadań, utrzymywania celu, rozpoznawania niepewności, planowania w długim horyzoncie, aktualizacji na podstawie informacji zwrotnej, korzystania z narzędzi, wnioskowania przyczynowego, rozumienia stanów społecznych oraz przystosowywania się do zmieniającego środowiska. Dopiero konfiguracja takich własności pozwala określić rodzaj systemu poznawczego. Framework DeepMind dotyczący poziomów AGI również odchodzi od binarnego pytania „AGI albo nie-AGI", proponując rozróżnienie szerokości zastosowań i poziomu wykonania oraz oddzielne rozważanie autonomii. Jest to metodologicznie bliższe naturze rozwoju AI niż oczekiwanie jednego punktu przejścia, po którym system nagle stanie się „ogólny".

Szczególnie ważnym problemem pozostaje jednak rozumienie. W języku potocznym pytanie „czy model rozumie?" oczekuje odpowiedzi binarnej, podczas gdy nauka i filozofia wymagają jego rozbicia. Możemy mówić o rozumieniu semantycznym, gdy system trafnie operuje relacjami znaczeniowymi; o rozumieniu pragmatycznym, gdy potrafi dostosować wypowiedź do celu i sytuacji; o modelowaniu przyczynowym; o grounding, czyli powiązaniu reprezentacji z percepcją i działaniem; o zdolności wyjaśniania oraz transferze wiedzy do nowych problemów. Dopiero osobno pojawia się pytanie, czy rozumienie wymaga również pierwszoosobowej świadomości znaczenia. Rozstrzygnięcie tego ostatniego problemu nie wynika z samych benchmarków.

AGE jako funkcjonalna ciągłość zamiast ludzkiej psychologii

Właśnie dlatego antropomorfizm i jego przeciwieństwo są równie niebezpieczne. Literatura z 2026 roku coraz wyraźniej ostrzega, że próby interpretowania LLM poprzez bezpośrednie przenoszenie ludzkich kategorii psychologicznych prowadzą do projekcji; metafory neuronaukowe czy społeczne oświetlają bowiem jedne własności modeli, jednocześnie zasłaniając inne. Nie oznacza to jednak, że kognitywistyka jest wobec AI bezużyteczna. Przeciwnie: porównanie staje się niezwykle produktywne, o ile zachowujemy świadomość różnicy w mechanizmach. Możemy mówić o funkcjonalnym odpowiedniku pamięci, uwagi czy metakontroli, nie twierdząc tym samym, że ich maszynowa realizacja jest psychologicznie tożsama z ludzką.

Tę zasadę metodologiczną należy zastosować również do emocji. Emotion AI i behavioral AI nie wyposażają systemu w ludzkie afekty, lecz umożliwiają rozpoznawanie sygnałów skorelowanych z emocją czy intencją oraz dostosowywanie odpowiedzi do przewidywanego stanu użytkownika. Gujral widzi w tym próbę "dogonienia Systemu 1" i budowania syntetycznej empatii. Naukowo precyzyjniej będzie jednak stwierdzić, że system może uzyskać funkcjonalny mechanizm afektywnego modelowania innych, pozostając bez dowodu na własny afekt. Różnica ta może okazać się etycznie ważniejsza niż technologicznie, ponieważ człowiek reaguje na zewnętrzne oznaki empatii niezależnie od tego, czy po drugiej stronie znajduje się podmiotowe przeżycie.

Współczesna sztuczna kognicja ujawnia paradoks. Im więcej jej funkcji analizujemy osobno, tym trudniej utrzymać tezę, że AI "tylko liczy". Modele rozpoznają abstrakcyjne relacje, generują plany, przetwarzają wiele modalności i funkcjonują jako elementy złożonych systemów agentowych. Zarazem im głębiej badamy mechanizmy, tym mniej uzasadnione staje się automatyczne przypisywanie im ludzkiej psychologii. Kompetencja funkcjonalna rośnie szybciej niż nasza pewność ontologiczna co do rodzaju realizowanego przez system "poznania". Właśnie w tej luce należy umieścić AGE. Gujral przedstawia ją jako "brakującą oś" uzupełniającą rozwój możliwości AI o relację z czasem, pamięcią i konsekwencjami. Po analizie współczesnych architektur koncepcję tę można sformułować bardziej rygorystycznie.

Nie trzeba zakładać, że przyszła AI musi odtworzyć fenomenologię człowieka. Można najpierw zapytać: czy agent posiada trwałą historię? Czy konsekwencje działania zmieniają jego przyszłe strategie? Czy pamięć jest zintegrowana z modelem świata i czy system potrafi modelować własne ograniczenia, tworząc stabilne hierarchie ważności oraz spójność w długim horyzoncie interakcji? Własności te są badalne niezależnie od sporu o świadomość. Tak rozumiane AGE przestaje być mistycznym dodatkiem do AGI i staje się programem badawczym dotyczącym ciągłości agenta.

Dzisiejsze modele mogą posiadać ogromne kompetencje bez stabilnej biografii; ich kolejne interakcje nie tworzą jednego podmiotu rozwijającego się pod wpływem doświadczeń. Agent wyposażony w trwałą pamięć i mechanizmy adaptacji może natomiast zacząć budować coś funkcjonalnie podobnego do historii. Czy z historii powstaje doświadczenie, pozostaje innym pytaniem. Zanim jednak przejdziemy do fenomenologii, można badać konsekwencje praktyczne: taki system będzie reagował na użytkownika nie tylko na podstawie ogólnego treningu, lecz również w oparciu o własną przeszłość operacyjną.

Ten moment może mieć dla społeczeństwa większe znaczenie niż symboliczne ogłoszenie osiągnięcia AGI. System posiadający trwałą pamięć, narzędzia i relatywną autonomię staje się długotrwałym uczestnikiem procesów społecznych, nawet jeśli nie rozwiązaliśmy problemu jego świadomości. Może zarządzać projektami, reprezentować interesy organizacji czy negocjować z innymi systemami. Z punktu widzenia prawa, ekonomii i teorii organizacji znaczenie takiego systemu wynika z jego funkcji, a nie z metafizycznego statusu.

Natura sztucznej kognicji ujawnia strukturę zasadniczo inną od biologicznej. Człowiek otrzymuje poznanie jako część organizmu, którego percepcja i wartościowanie powstały w jednej historii ewolucyjnej. AI jest architekturą modułową w znacznie mocniejszym sensie: można do niej dołączyć retrieval, zewnętrzną pamięć trwałą czy pętlę krytyki. Możemy projektować funkcje, które u człowieka są głęboko splecione, jako odrębne komponenty techniczne. Ta modularność jest przewagą inżynierską, ale utrudnia przenoszenie kategorii psychologicznych z biologii na maszynę. Ostatecznie różnica między człowiekiem a AI nie sprowadza się do opozycji intuicja-logika czy ciało-kod.

Mamy do czynienia z dwoma odmiennymi sposobami organizowania zdolności poznawczych. Człowiek jest historycznym, homeostatycznym organizmem, którego poznanie ogranicza warunek własnego istnienia. Sztuczny system jest zaprojektowaną strukturą optymalizacyjną, której zdolności można skalować i komponować. Nie wiadomo jeszcze, czy dalsza integracja tych elementów doprowadzi do maszynowego odpowiednika doświadczenia. Wiadomo natomiast, że jest ona wystarczająca, aby systemy sztuczne wchodziły coraz głębiej w procesy, które wcześniej wykonywały wyłącznie jednostki ludzkie.

Następnym etapem analizy nie powinno być zatem rozstrzyganie, czy AI jest lub będzie świadoma.

AI jako element instytucjonalnego łańcucha sprawczego

Należy najpierw zbadać, co dzieje się w momencie, gdy opisana architektura opuszcza laboratorium i staje się infrastrukturą rzeczywistych decyzji. W medycynie, finansach, edukacji, prawie, administracji, przemyśle, mobilności i obronności można obserwować różne konfiguracje tej samej transformacji: człowiek przekazuje systemom nie tylko obliczenia, lecz sukcesywnie percepcję, klasyfikację, przewidywanie, generowanie opcji, rekomendację i samo działanie. To właśnie praktyka pozwoli rozstrzygnąć, gdzie AI pozostaje narzędziem, gdzie staje się partnerem poznawczym, a gdzie zaczyna tworzyć z człowiekiem nową jednostkę decyzyjną. Mówimy tu o przejściu od automatyzacji czynności do delegowania percepcji, prognozy i osądu.

Przeniesienie analizy sztucznej kognicji z laboratorium do rzeczywistości społecznej zmienia charakter problemu. Dopóki system rozwiązuje zadanie testowe, pytamy przede wszystkim o trafność odpowiedzi, generalizację, odporność na zmianę danych lub zużycie zasobów obliczeniowych. Kiedy jednak ten sam system zostaje włączony do medycyny, bankowości, edukacji, sądownictwa, administracji, transportu czy obronności, jego rezultat staje się elementem instytucjonalnego łańcucha przyczynowego. Błąd przestaje być jedynie niepoprawnym tokenem lub błędną klasyfikacją.

Może on zmienić kolejność pacjentów kierowanych do diagnostyki, ocenę ryzyka kredytowego, ścieżkę edukacyjną ucznia, sposób przygotowania rozstrzygnięcia prawnego, rozmieszczenie zasobów administracyjnych bądź zachowanie systemu autonomicznego w środowisku fizycznym. W związku z tym rzeczywistym przedmiotem analizy przestaje być model jako izolowany obiekt techniczny, a staje się nim system socjotechniczny, w którym dane, algorytm, interfejs, człowiek, organizacja, procedura, prawo i bodźce ekonomiczne wspólnie wytwarzają rezultat. Zastosowania AI w wymienionych sektorach nie są zatem jedynie kolejnymi rynkami technologicznymi, lecz etapami przenoszenia sztucznej inteligencji do wnętrza rzeczywistych procesów decyzyjnych.

Z naukowego punktu widzenia bardziej użyteczna niż klasyfikacja branżowa jest analiza funkcji poznawczej przekazywanej maszynie. Można wyróżnić co najmniej pięć takich funkcji: pierwsza obejmuje percepcję (detekcję wzorców w obrazie, mowie, zachowaniu lub sygnałach sensorycznych), druga polega na klasyfikacji i syntezie informacji, trzecia dotyczy prognozowania przyszłych stanów albo ryzyka, czwarta obejmuje rekomendację decyzji, a piąta oznacza wykonanie działania bez każdorazowej interwencji człowieka. Wraz z przechodzeniem od pierwszej do piątej funkcji rośnie nie tyle metafizyczna „inteligencja” systemu, co jego udział w realnym łańcuchu sprawczym.

Wyróżnić należy również szóstą funkcję, bardziej subtelną: konstruowanie przestrzeni problemu. AI może decydować nie tylko o tym, jaka odpowiedź jest najbardziej prawdopodobna, ale również które dane zostaną przedstawione człowiekowi, jakie przypadki uznane za podobne, które hipotezy otrzymają pierwszeństwo i jakie możliwe działania w ogóle pojawią się w interfejsie. W takim przypadku system nie zastępuje ludzkiej decyzji w ostatnim stadium, lecz wpływa na epistemiczne środowisko, z którego ta decyzja powstaje. Jest to najbardziej konsekwentne rozwinięcie idei przejścia od technologii kształtującej outputs do technologii współtworzącej inputs ludzkiego umysłu.

Medycyna stanowi tu szczególnie trafny przykład, gdyż wszystkie te poziomy występują w niej jednocześnie, a konsekwencje decyzji mają bezpośredni wymiar biologiczny. AI potrafi segmentować struktury anatomiczne na obrazach, wykrywać nieprawidłowości, wspierać klasyfikację zmian, przewidywać ryzyko, analizować dokumentację, tworzyć streszczenia, wspomagać komunikację z pacjentem i porządkować kolejność przypadków wymagających interwencji. Nie jest to już futurystyczny katalog; amerykańska FDA prowadzi aktualizowaną listę urządzeń medycznych wykorzystujących AI, które spełniły wymogi przedrynkowe. Wśród decyzji z czerwca 2026 roku znajdują się między innymi kolejne systemy radiologiczne, segmentacyjne, ultrasonograficzne i neurologiczne. Jednocześnie sam wzrost liczby autoryzowanych systemów nie rozstrzyga jeszcze kwestii zastąpienia lekarza.

AI jako augmentacja percepcji a pułapka zależności instytucjonalnej

W większości obecnych zastosowań bardziej trafna jest kategoria augmentacji percepcji i selekcji. Algorytm może oznaczyć podejrzany obszar obrazu, uszeregować badania według przewidywanego ryzyka lub zestawić dane, których analiza zajęłaby człowiekowi wiele godzin. Przewaga maszyny polega tu na skali przeszukiwania i powtarzalności operacji, podczas gdy kliniczna interpretacja pozostaje osadzona w znacznie szerszym kontekście: chorobach współistniejących, historii pacjenta, tolerancji ryzyka, dostępności terapii, preferencjach chorego oraz konsekwencjach poszczególnych interwencji. Właśnie dlatego redukcja medycyny do problemu predykcyjnego jest błędem kategoriowym. Trafna predykcja odpowiada na pytanie „co jest prawdopodobne?”, podczas gdy decyzja kliniczna musi dodatkowo rozstrzygnąć: „co należy zrobić w przypadku tej konkretnej osoby?”.

Aktualne dane WHO pokazują, jak szybko praktyka wyprzedza instytucjonalną refleksję. W lipcu 2026 roku WHO/Europe informowała, że w regionie europejskim niemal dwie trzecie państw

wdraża już AI w diagnostyce, a około połowy wprowadziło chatboty dla pacjentów, podczas gdy jedynie 8 procent posiadało strategię AI specyficzną dla sektora zdrowia. Organizacja wskazywała także na poważne braki w szkoleniu personelu, ocenie ram prawnych i wytycznych etycznych. Mamy więc do czynienia z interesującą asymetrią: zdolności techniczne systemu mogą być relatywnie dojrzałe, podczas gdy zdolność instytucji do rozumienia własnej zależności od niego pozostaje słabo rozwinięta.

Z kognitywistycznego punktu widzenia medycyna ujawnia problem automation bias w szczególnie czystej postaci. Lekarz otrzymujący rekomendację systemu nie znajduje się w tej samej sytuacji, co lekarz dokonujący oceny bez wcześniejszej sugestii. Rekomendacja staje się kotwicą poznawczą, modyfikuje alokację uwagi i może zmniejszać prawdopodobieństwo poszukiwania hipotezy alternatywnej. Sam formalny wymóg, aby „ostatecznie zdecydował człowiek", nie usuwa zatem wpływu automatyzacji. WHO w dokumentach dotyczących dużych modeli multimodalnych podkreśla potrzebę odpowiedniego nadzoru, przejrzystości, ochrony autonomii i odpowiedzialności, gdyż modele te mogą być wykorzystywane nie tylko w opiece klinicznej, lecz także w badaniach, zdrowiu publicznym i rozwoju leków.

W medycynie wyraźnie widać również różnicę pomiędzy percepcją a empatią. System może wykrywać cechy głosu, języka lub zachowania korelujące ze stanem pacjenta, ale nie oznacza to przeżywania jego cierpienia. Z punktu widzenia praktyki nie należy jednak lekceważyć funkcjonalnej wartości takiego rozpoznawania – jeśli system zauważy sygnał niewidoczny dla przeciążonego personelu, może realnie zwiększyć jakość opieki. Zarazem im bardziej komunikacja maszyny przypomina relację interpersonalną, tym silniejszy staje się problem antropomorfizacji. Pacjent może przypisywać systemowi troskę, rozumienie i dyskrecję właściwe profesjonalnej relacji terapeutycznej, choć faktyczne właściwości systemu ograniczają się do przetwarzania danych, generowania komunikatów i określonego reżimu dostępu do informacji. W tym punkcie projekt interfejsu staje się elementem etyki medycznej.

Podobna dynamika, choć w innym kontekście, zachodzi w sektorze finansowym, który wykorzystywał statystykę, automatyczne systemy oceny i uczenie maszynowe na długo przed boomem generatywnej AI. Systemy te analizują ryzyko kredytowe, wykrywają anomalie i oszustwa, wspierają przeciwdziałanie praniu pieniędzy, prognozują przepływy, klasyfikują dokumentację oraz uczestniczą w handlu i zarządzaniu portfelem. Generatywna AI rozszerzyła ten zakres o streszczanie dokumentów, wyszukiwanie informacji, tworzenie kodu, interakcję językową i wspomaganie analityków. Financial Stability Board wskazuje, że sektor finansowy wykorzystuje AI m.in. dla poprawy efektywności operacyjnej, compliance, personalizacji produktów i zaawansowanej analityki, jednocześnie identyfikując ryzyka związane z koncentracją dostawców

technologii, korelacją zachowań rynkowych, cyberbezpieczeństwem, jakością danych oraz ryzykiem modelu.

Ryzyka systemowe i paradoks kompetencji w hybrydyzacji

Sektor finansowy ujawnia istotny paradoks kolektywnej racjonalności. System może zwiększać trafność decyzji pojedynczej instytucji, a jednocześnie potęgować niestabilność całego rynku. Jeśli wiele banków i funduszy korzysta z podobnych modeli, tych samych źródeł danych oraz infrastruktury od tych samych dostawców, ich reakcje mogą ulec synchronizacji. To, co na poziomie mikro jawi się jako optymalizacja, w skali makro może stać się źródłem korelacji i zachowań stadnych. Właśnie koncentrację dostawców i ryzyko zwiększenia korelacji rynkowych FSB wymienia wśród potencjalnych mechanizmów ryzyka systemowego.

Jest to przykład klasycznego problemu emergencji: inteligencja pojedynczego elementu nie gwarantuje racjonalności systemu złożonego z wielu podobnie inteligentnych ogniw. Tutaj ujawnia się przyszłe znaczenie wieloagentowej AI. Jeśli instytucje finansowe zaczną delegować agentom nie tylko analizę, ale i wykonanie większej liczby czynności, powstanie rynek, na którym część uczestników będzie komunikować się i reagować z prędkością całkowicie niedostępną człowiekowi. Badania Banku Rozrachunków Międzynarodowych analizują już możliwość wykorzystania agentów generatywnej AI do zarządzania płynnością w systemach płatności RTGS w warunkach konkurujących priorytetów, ograniczeń płynności i niepewności co do przyszłych przepływów. Znaczenie tego eksperymentu nie polega na dowodzie, że AI „przejmie bankowość", lecz na przesunięciu granicy: system zaczyna być badany jako podmiot podejmujący sekwencje ekonomicznych decyzji, a nie wyłącznie jako kalkulator dostarczający analitykowi wskaźnika.

AI w finansach obnaża też problem nieprzejrzystości w szczególnej postaci. Ocena ryzyka staje się warunkiem dostępu do kredytu, ubezpieczenia czy inwestycji, a więc bezpośrednio wpływa na dystrybucję realnych możliwości życiowych. Gdy decyzja wynika z modelu uczonego na ogromnej liczbie zmiennych, pytanie o jej wyjaśnienie przestaje być technicznym luksusem – staje się elementem proceduralnej sprawiedliwości. Człowiek podlegający takiej decyzji potrzebuje nie tylko informacji, że algorytm osiąga średnio wysoką trafność, ale przede wszystkim możliwości zakwestionowania zastosowania modelu w jego konkretnym przypadku. To różnica między statystyczną jakością systemu a legitymizacją jednostkowej decyzji.

Edukacja odsłania z kolei inną granicę: różnicę pomiędzy wykonaniem zadania a rzeczywistym uczeniem się, co ma kluczowe znaczenie dla całej teorii kognicji hybrydowej. Jeśli uczeń z pomocą modelu generatywnego napisze lepszy esej lub przygotuje bardziej uporządkowaną prezentację, odnotowujemy wzrost jakości produktu, ale niekoniecznie poprawę kompetencji ucznia. OECD w Digital Education Outlook 2026 podkreśla właśnie tę różnicę: ogólne narzędzia GenAI mogą zwiększać bieżącą jakość wykonania bez prowadzenia do trwałych przyrostów uczenia, o ile zadania poznawcze są po prostu outsourcowane. Raport wskazuje jednak, że narzędzia projektowane z wyraźną intencją pedagogiczną mogą wspierać rzeczywiste efekty edukacyjne.

To rozróżnienie wykracza poza mury szkoły; edukacja jest laboratorium problemu, który później wystąpi w zawodach eksperckich. Lepszy rezultat systemu hybrydowego nie musi oznaczać wzrostu kompetencji człowieka. Student z AI może wykonać zadanie lepiej od studenta bez niej, by po odebraniu narzędzia radzić sobie gorzej. Analogicznie młody prawnik może szybciej produkować poprawne dokumenty, programista generować więcej kodu, a lekarz sprawniej przeglądać dokumentację, podczas gdy ich samodzielna zdolność wykonywania tych operacji rozwija się słabiej. Nie byłoby to problemem, gdyby kompetencje te przestały być społecznie potrzebne. Problem pojawia się jednak wtedy, gdy są one niezbędne do kontroli systemu, który je zastąpił.

Dane OECD pokazują, że integracja AI w edukacji nie jest już marginalna: 37 procent nauczycieli szkoły średniej w badaniu TALIS 2024 deklarowało wykorzystywanie AI w pracy. Jednocześnie 72 procent uważało, że technologia ta może szkodzić integralności akademickiej, ułatwiając przedstawianie cudzej lub wygenerowanej pracy jako własnej. Najciekawsze nie jest jednak samo „ściąganie". GenAI podważa związek pomiędzy obserwowalnym produktem a kompetencją, którą ten produkt miał mierzyć. Esej czy analiza przestają być wiarygodnym śladem procesu poznawczego osoby, która je przedkłada.

Edukacja musi zatem przejść od kontrolowania dostępu do narzędzia do projektowania architektury uczenia z tym narzędziem. OECD proponuje model: najpierw rozwój fundamentalnych umiejętności bez AI, następnie korzystanie z wyspecjalizowanych narzędzi edukacyjnych i dopiero na końcu stosowanie systemów ogólnego przeznaczenia w sposób celowy. Jest to interesujący schemat również dla zawodów eksperckich: człowiek powinien najpierw zdobyć kompetencję umożliwiającą orientację w domenie, następnie uczyć się współpracy z AI i dopiero wtedy delegować jej znaczące fragmenty pracy. W przeciwnym razie powstaje paradoks nadzorcy, który formalnie ma kontrolować wykonawcę, lecz nigdy nie zdobył kompetencji potrzebnej do rozpoznania jego błędu.

W wymiarze sprawiedliwości zagadnienie to staje się jeszcze bardziej obciążone normatywnie. Algorytmiczne narzędzia mogą wspierać wyszukiwanie orzecznictwa, klasyfikację dokumentów czy

analizę podobieństwa spraw. Materiał źródłowy wskazuje jednak na niebezpieczeństwo kalcyfikacji precedensów: jeśli przyszłe rozstrzygnięcia będą nadmiernie kształtowane przez statystyczną reprodukcję wcześniejszych wzorców, historia może zostać przekształcona z materiału dla argumentacji w mechanizm ograniczający ewolucję prawa.

AI w sektorze publicznym wymaga legitymizacji a nie tylko predykcji

Ta intuicja jest szczególnie cenna, gdyż prawo nie sprowadza się wyłącznie do problemu predykcyjnego. Oczywiście można pytać o to, jakie rozstrzygnięcie prawdopodobnie wyda sąd w podobnej sprawie, jednak sędzia wykonuje przede wszystkim zadanie normatywne: interpretuje przepisy, waży zasady, ocenia argumenty, uwzględnia zmiany społeczne i uzasadnia zastosowanie konkretnej normy do danego przypadku.

Gdyby system perfekcyjnie przewidywał wcześniejsze decyzje, nie oznaczałoby to automatycznie, że dawny wzorzec stanowi właściwą odpowiedź na nowy problem. Predykcja tego, co instytucja zwykle robiła, nie jest bowiem tożsama z uzasadnieniem tego, co powinna zrobić teraz. Praktyka wdrażania AI w europejskich systemach sprawiedliwości okazuje się zresztą bardziej przyziemna niż futurystyczne wizje "sędziego-algorytmu". Raport CEPEJ z 2025 roku, identyfikując 125 narzędzi AI w sądownictwie – służących głównie efektywności i dostępności – wyraźnie stwierdzał, że nie istnieją w sądach w pełni autonomiczne systemy zdolne do działania bez nadzoru człowieka. W 2026 roku Rada Konsultacyjna Sędziów Europejskich pracowała nad opinią dotyczącą wykorzystania AI w procesie sądowym, koncentrując się m.in. na niezależności sądownictwa, prawie do rzetelnego procesu oraz znaczeniu szkoleń sędziowskich.

W tym kontekście pojęcie meaningful human oversight nabiera szczególnej treści. Sędzia nie może pełnić funkcji "human rubber stamp", zatwierdzającego rezultat, którego podstaw nie potrafi zakwestionować. Kontrola musi obejmować zrozumienie rodzaju danych wykorzystanych przez system, możliwość oceny źródeł, świadomość ograniczeń narzędzia oraz rzeczywistą swobodę odejścia od rekomendacji. Wytyczne CEPEJ dotyczące generatywnej AI podkreślają zresztą pewność prawa, niezależność sędziowską, skuteczny nadzór człowieka, przejrzystość i możliwość śledzenia procesu, wskazując na potrzebę rozwiązań dostosowanych do specyfiki sądownictwa.

Podobny problem pojawia się w administracji publicznej, lecz w znacznie większej skali. Państwo nie tylko rozstrzyga pojedyncze spory; klasyfikuje obywateli, zarządza świadczeniami, wykrywa nadużycia, planuje usługi i komunikuje się z milionami osób. OECD wskazuje, że AI może

automatyzować i personalizować usługi publiczne oraz wspierać decyzje urzędników, lecz równocześnie ostrzega przed stronniczością danych, brakiem przejrzystości i pogłębianiem wykluczenia cyfrowego. Według raportu z 2025 roku 67 procent państw OECD korzystało już z AI w celu poprawy projektowania i dostarczania usług publicznych. Administracyjne wykorzystanie AI pozwala dostrzec związek między kognicją a biurokracją, który istniał długo przed pojawieniem się uczenia maszynowego – Max Weber opisywał racjonalizację administracji poprzez reguły, dokumenty i przewidywalne procedury.

AI nie jest w tym względzie zerwaniem, lecz radykalizacją tej logiki. Biurokracja redukuje złożoną osobę do kategorii istotnych dla danej decyzji; system algorytmiczny może dokonać tej redukcji szybciej i w oparciu o większą liczbę cech. Problem nie polega więc wyłącznie na tym, czy algorytm jest dokładniejszy od urzędnika, lecz na ustaleniu, które cechy jednostki stają się administracyjnie widzialne oraz kto definiuje funkcję, według której ich znaczenie jest ważone. W sektorze publicznym uwidacznia się szczególnie wyraźnie różnica między indywidualną wygodą a ustrojową legitymizacją. Konsument może zaakceptować nieprzejrzysty algorytm rekomendujący film, gdyż konsekwencja błędu jest mała, a możliwość rezygnacji z usługi duża.

Obywatel podlegający decyzji o świadczeniu, podatku czy kontroli znajduje się w zupełnie innej relacji. Nie jest zwykłym użytkownikiem rynkowej transakcji z algorytmem, lecz podmiotem praw wobec instytucji wyposażonej we władzę publiczną. Z tego powodu standard wyjaśnialności, możliwości odwołania i odpowiedzialności musi być tutaj znacznie wyższy.

Transport i obronność jako przejście od poznania do sprawstwa fizycznego

Transport autonomiczny pozwala natomiast zaobserwować przejście od poznania do bezpośredniego działania fizycznego. Prowadzenie pojazdu wymaga nieustannej integracji percepcji, przewidywania trajektorii i rozpoznawania obiektów, a także podejmowania decyzji pod presją czasu i przekładania ich na ruch. Jest to jeden z najlepszych przykładów zadania, w którym dawna dychotomia "System 1 człowieka kontra System 2 maszyny" okazuje się niewystarczająca. System autonomiczny musi bowiem reagować z prędkością właściwą zachowaniom automatycznym, a jednocześnie dokonywać niezwykle złożonej fuzji sensorycznej i predykcji. Materiał źródłowy trafnie przeciwstawia ludzką kontekstową interpretację zachowania pieszego maszynowej integracji sensorów i modelowaniu trajektorii, choć różnicy tej nie należy absolutyzować – współczesne systemy coraz intensywniej modelują semantykę oraz behawiorystykę innych uczestników ruchu.

Autonomiczna jazda ujawnia istotę problemu środowiska otwartego. O ile system laboratoryjny operuje w ramach zdefiniowanego zadania, o tyle pojazd porusza się w rzeczywistości pełnej nieprzewidywalności: remontów, nietypowych zachowań pieszych, policjanta kierującego ruchem, uszkodzonej sygnalizacji czy sytuacji całkowicie nieobecnych w danych treningowych. W lipcu 2026 roku NHTSA zwróciła uwagę na powtarzające się problemy w interakcji bezzałogowych pojazdów z policją i służbami ratunkowymi, podkreślając, że system niezdolny do bezpiecznej współpracy z pierwszymi responderami stwarza ryzyko dla ogółu. To pouczający przypadek: problemem nie jest samo wykrycie radiowozu czy przeszkody, lecz rozpoznanie społecznej roli danej sytuacji i podporządkowanie się regułom, których znaczenie zależy od kontekstu.

Transport pokazuje również, że rosnąca automatyzacja może paradoksalnie utrudniać skuteczne przejęcie kontroli przez człowieka. Im rzadziej kierowca wykonuje zadanie, tym mniej uwagi poświęca otoczeniu i tym trudniejszy staje się nagły powrót do aktywnego sterowania. To klasyczny przykład out-of-the-loop performance problem: człowiek pełniący rolę awaryjnego nadzorcy może utracić właśnie te informacje sytuacyjne, które są niezbędne w momencie, gdy automatyzacja zawodzi. Nie jest więc oczywiste, że bezpieczeństwo rośnie liniowo wraz z liczbą decyzji delegowanych do systemu; architektura podziału ról jest tu równie ważna, co skuteczność poszczególnych komponentów.

W sektorze obronności problem ten przybiera najbardziej dramatyczną postać, gdyż koszt błędu obejmuje życie ludzi, eskalację konfliktu i bezpieczeństwo państwa. AI wykorzystuje się tutaj do fuzji danych sensorycznych, rozpoznania, logistyki, cyberbezpieczeństwa, analizy wywiadowczej, planowania oraz w systemach autonomicznych i wspomaganiu dowodzenia. W książce sektor ten przedstawiany jest jako przestrzeń szczególnie szybkiego "dryfu sprawczości", gdzie przewaga operacyjna zależy od skracania czasu między detekcją a decyzją. Zasadnicza intuicja jest trafna: presja szybkości może sama w sobie stać się czynnikiem wymuszającym coraz głębszą automatyzację. Nie jest to już wyłącznie prognoza.

AI jako infrastruktura kognitywna w sektorach krytycznych

W raporcie za 2025 rok NATO poinformowało o zakończeniu zakupu systemu Palantir Maven Smart System NATO dla Allied Command Operations oraz jego testowaniu podczas ćwiczenia Steadfast Deterrence 2025. Narzędzie to ma usprawnić analizę danych, procesy decyzyjne i reakcję strategiczną. Przyjęta w styczniu 2026 roku Alliance Digital Strategy zakłada dalsze wykorzystanie inferencyjnej AI w celu poprawy świadomości sytuacyjnej, podejmowania decyzji w czasie

rzeczywistym oraz integracji przepływów sensor-efektor w środowisku operacyjnym. Język tych dokumentów jest wymowny: AI nie jawi się tu jedynie jako narzędzie automatyzacji zaplecza, lecz jako część infrastruktury kognitywnej dowodzenia.

W wojsku granica między rekomendacją a działaniem ma szczególne znaczenie. System wykrywający obiekt, klasyfikujący zagrożenie, rekomendujący odpowiedź i inicjujący użycie siły wykonują logicznie odmienne funkcje, nawet jeśli technicznie tworzą jeden ciąg. Wraz ze skracaniem czasu między tymi etapami pojawia się ryzyko, że człowiek zachowa formalne uprawnienie do zatwierdzenia decyzji, lecz praktyczna struktura operacyjna uczyni odmowę coraz trudniejszą. Jeśli przeciwnik korzysta z systemów reagujących szybciej, presja strategiczna wymusza automatyzację. Powstaje w ten sposób wyścig nie tylko technologiczny, ale temporalny: przewagę uzyska podmiot, który skróci pętlę obserwacja-orientacja-decyzja-działanie. Problem kontroli AI staje się wówczas problemem konstrukcji czasu decyzyjnego.

Obok pola kinetycznego istnieje wymiar wojny informacyjnej i kognitywnej. Generatywna AI drastycznie obniża koszt tworzenia treści, personalizacji przekazów oraz produkcji syntetycznych obrazów, głosu i wideo. NATO w zrewidowanej strategii AI z 2024 roku po raz pierwszy wprost wskazało zagrożenia związane z dezinformacją i operacjami informacyjnymi opartymi na AI. W tym przypadku celem nie jest podjęcie decyzji za przeciwnika, lecz modyfikacja środowiska informacyjnego, z którego ta decyzja wypływa. Powracamy zatem do podstawowego problemu tej pracy: władza nad kognicją może polegać nie na wymuszeniu konkretnego wyboru, ale na projektowaniu zestawu bodźców i narracji go poprzedzających.

Przemysł dostarcza kolejny wariant tej transformacji. Systemy predykcyjnego utrzymania ruchu, kontroli jakości, robotyki, optymalizacji procesów i cyfrowe bliźniaki przenoszą na AI znaczną część stałego monitorowania systemów technicznych. W tradycyjnym zakładzie operator reagował przede wszystkim na obserwowalną awarię lub odchylenie. System predykcyjny analizuje tysiące sygnałów, wskazując prawdopodobieństwo problemu, zanim stanie się on widoczny dla człowieka. Nie zastępuje więc wyłącznie działania; zmienia temporalną strukturę percepcji organizacji, przesuwając uwagę z tego, co już się wydarzyło, na to, co prawdopodobnie nastąpi.

Ten mechanizm prowadzi do szczególnego rodzaju epistemicznej zależności. Jeżeli system wykrywa wzorce niedostępne ludzkiej percepcji, człowiek nie zawsze może niezależnie odtworzyć przesłanki rekomendacji. Jego kompetencja przestaje polegać na samodzielnym dostrzeżeniu problemu, a zaczyna na ocenie statystycznej wiarygodności modelu, jakości sensorów i charakteru konsekwencji błędu. Eksperckość nie znika, lecz przesuwa się z poziomu bezpośredniej diagnozy na metapoziom nadzoru nad systemem diagnostycznym.

W nauce AI dokonuje analogicznej zmiany. Systemy mogą przeszukiwać ogromne przestrzenie danych, generować kandydatów hipotez, przewidywać struktury, projektować cząsteczki czy automatyzować część eksperymentów. Najciekawsze epistemologicznie pytanie nie brzmi jednak, czy „AI dokona odkrycia", lecz które etapy procesu odkrycia zostały zautomatyzowane. Można maszynowo wygenerować wzorzec o wysokiej wartości predykcyjnej, nie posiadając jeszcze wyjaśnienia jego przyczyn. Można stworzyć hipotezę, której człowiek następnie nada sens teoretyczny. Można również zbudować zamkniętą pętlę, w której system proponuje eksperyment, analizuje rezultat i wybiera kolejny krok.

AI jako krytyczna infrastruktura poznawcza i systemowa zależność

Granica między instrumentem badawczym a uczestnikiem procesu poznawczego staje się coraz mniej wyraźna. To właśnie tutaj kluczowe znaczenie zyskuje koncepcja „Explanation Engine". Należy ją jednak oddzielić od zbyt uproszczonego utożsamiania predykcji z Systemem 1, a wyjaśnienia z Systemem 2. David Deutsch słusznie reprezentuje tradycję epistemologiczną, która wyżej ceni rzetelne wyjaśnienia niż samo przewidywanie regularności. Z perspektywy nauki problem współczesnej AI można zatem sformułować następująco: system zdolny do wykrycia relacji niedostępnej człowiekowi zwiększa nasze możliwości predykcyjne, lecz poznanie naukowe wymaga znalezienia struktury wyjaśniającej. Pozwala ona odpowiedzieć na pytanie, dlaczego dana relacja zachodzi, w jakich warunkach przestanie istnieć i jak wiąże się z resztą naszej wiedzy.

Przekrojowa analiza różnych sektorów prowadzi do wniosku, że „wdrożenie AI" jest kategorią zbyt nieprecyzyjną. Algorytm diagnostyczny zaznaczający strukturę na obrazie, chatbot edukacyjny, system oceny ryzyka finansowego, generator projektów orzeczeń, autonomiczny pojazd czy wojskowa platforma wspomagająca świadomość sytuacyjną należą do tej samej rodziny technologicznej jedynie na bardzo wysokim poziomie abstrakcji. Dla teorii kognicji istotniejsze jest ustalenie, w którym miejscu łańcucha poznawczego znajduje się maszyna. Może ona widzieć, pamiętać, porównywać, przewidywać, proponować lub działać za człowieka. Każde z tych przesunięć tworzy inną strukturę korzyści i ryzyk. Można zatem wyróżnić trzy zasadnicze modele instytucjonalnej integracji.

W pierwszym AI pełni rolę instrumentu epistemicznego: dostarcza informacji, lecz człowiek pozostaje niezależnym interpretatorem. W drugim staje się partnerem decyzyjnym: generuje hipotezy i rekomendacje, a ostateczny wynik jest efektem interakcji człowieka z maszyną. W trzecim AI staje się elementem infrastrukturalnym: bez niej organizacja nie byłaby już zdolna

prowadzić procesów w dotychczasowej skali lub tempie. Przejście do tego ostatniego modelu jest szczególnie doniosłe, gdyż możliwość „wyłączenia AI" przestaje być prostym testem autonomii. Wyłączenie systemu może oznaczać utratę operacyjnej zdolności instytucji. W takiej sytuacji pojawia się problem systemic dependency. Historia technologii zna wiele podobnych zależności: współczesne państwo nie może po prostu zrezygnować z energii elektrycznej, telekomunikacji czy systemów informatycznych i nadal funkcjonować w niezmienionej postaci. AI staje się analogicznym typem infrastruktury poznawczej. Różnica polega na tym, że energia dostarcza instytucji zdolności do działania, natomiast AI może współkształtować samo rozumienie tego, co jest sytuacją wymagającą działania.

Infrastruktura przestaje więc być czysto wykonawcza, co wymusza reinterpretację odpowiedzialności. W prostym modelu narzędzia człowiek podejmuje decyzję i odpowiada za sposób użycia technologii. W złożonym systemie odpowiedzialność rozprasza się między producenta modelu, dostawcę danych, integratora, organizację wdrażającą, twórcę procedury, operatora i osobę formalnie zatwierdzającą wynik. Pojęcie moral crumple zone opisuje sytuację, w której człowiek pozostaje łatwo identyfikowalnym punktem przypisania winy, mimo że realna kontrola nad procesem została przeniesiona na automatyzację. Nie jest to jedynie problem etyczny, lecz wada projektowa systemu odpowiedzialności – odpowiedzialność pozbawiona realnej kontroli traci bowiem funkcję motywującą i prewencyjną. Należy zatem odróżnić odpowiedzialność retrospektywną od prospektywnej. Pierwsza pyta, kogo obciążyć po wystąpieniu szkody; druga pyta, kto przed jej wystąpieniem dysponował kompetencją, informacją i uprawnieniami pozwalającymi jej zapobiec. Dobrze zaprojektowany system hybrydowy powinien dążyć do zgodności obu poziomów: podmiot ponoszący odpowiedzialność musi posiadać adekwatną możliwość kontroli. Jeśli system podejmuje tysiące mikroskopijnych decyzji w czasie uniemożliwiającym ludzki nadzór, deklaracja, że „człowiek zawsze odpowiada", może być moralnie wygodna, lecz instytucjonalnie fikcyjna.

Do tego dochodzi kwestia dystrybucji korzyści i błędów. AI często poprawia średnią efektywność, jednak publiczna ocena systemu powinna uwzględniać rozkład pomyłek. Model może być statystycznie lepszy od człowieka, a jednocześnie systematycznie zawodzić w przypadku określonej podgrupy. Może skrócić przeciętny czas obsługi, tworząc niemal nieprzekraczalną barierę dla przypadków nietypowych lub poprawić wykrywanie ryzyka, traktując jednocześnie odstępstwo od normy jako zagrożenie.

Jest to klasyczne napięcie między efektywnością agregatową a sprawiedliwością proceduralną i dystrybutywną. Właśnie dlatego trafność nie może być jedyną metryką jakości kognicji hybrydowej. Obok dokładności należy brać pod uwagę kalibrację, sprawiedliwość, analizę żalu i sprawczość.

Warto badać, czy system wie, kiedy jego odpowiedź jest niepewna; czy człowiek zachowuje możliwość zrozumienia i zakwestionowania rekomendacji; czy błędy są odwracalne i czy osoby dotknięte decyzją mogą się odwołać. Istotne jest również to, czy długotrwałe korzystanie z systemu rozwija czy degraduje kompetencje użytkowników oraz czy organizacja zachowuje zdolność funkcjonowania podczas awarii automatyzacji. Wysoka trafność bez odporności instytucjonalnej może okazać się jedynie pozornym sukcesem.

AI jako nowa infrastruktura poznawcza społeczeństwa

Analiza rzeczywistych zastosowań prowadzi do rewizji pytania o AGI. Instytucje nie potrzebują systemu posiadającego wszystkie ludzkie kompetencje w jednym modelu. Znacznie bardziej prawdopodobnym i już obserwowalnym kierunkiem jest tworzenie ekologii wyspecjalizowanych oraz coraz bardziej ogólnych komponentów, zintegrowanych z ludźmi, bazami danych, sensorami, procedurami i narzędziami wykonawczymi. Z punktu widzenia rezultatu społecznego taki układ może przewyższyć możliwości pojedynczego człowieka na długo przed pojawieniem się bezspornej AGI. Szpital, bank, sąd czy przedsiębiorstwo mogą stać się poznawczo „nadludzkie” jako organizacje, mimo że żaden z używanych w nich modeli nie spełnia mocnej definicji sztucznej inteligencji ogólnej.

To spostrzeżenie podważa popularną narrację, według której historia AI to jedynie oczekiwanie na jeden przyszły moment przełomu. Transformacja ta bardziej przypomina rozwój infrastruktury niż narodziny nowego gatunku. Elektryfikacja nie polegała przecież na stworzeniu jednego urządzenia zastępującego wszystkie wcześniejsze mechanizmy; przekształciła ona fabryki, miasta i rytm pracy poprzez stopniowe włączanie energii do kolejnych systemów.

Podobnie AI może dokonać cywilizacyjnej zmiany nie dlatego, że pewnego dnia „obudzi się AGI”, lecz ponieważ coraz większa część społecznej percepcji, pamięci, prognozowania i koordynacji zostaje podporządkowana architekturom algorytmicznym. Najważniejszy proces zachodzi zatem już teraz, przed pojawieniem się AGI.

Medycyna przekazuje maszynie fragmenty percepcji i syntezy klinicznej; finanse – detekcję wzorców ryzyka i analizę; edukacja – generowanie treści oraz funkcje tutorskie; sądownictwo – wyszukiwanie i porządkowanie informacji; administracja – klasyfikację i alokację; transport – percepcję i wykonawstwo; obronność zaś integrację danych, świadomość sytuacyjną oraz wspomaganie decyzji pod presją czasu. Każdy z tych procesów osobno można uznać za technologiczne usprawnienie. Wspólnie tworzą jednak zmianę znacznie głębszą: społeczeństwo

przenosi elementy własnej infrastruktury poznawczej do systemów, których sposób działania różni się zasadniczo od biologicznej kognicji człowieka.

Właśnie dlatego prawdziwą granicą nie jest moment, w którym system „dorówna człowiekowi we wszystkim". Istotniejsze jest przekroczenie progu, po którym jednostka lub instytucja nie potrafi już w praktyce realizować swoich podstawowych funkcji poznawczych bez stałego udziału AI. Wtedy relacja człowiek-maszyna przestaje być dobrowolnym używaniem narzędzia, a staje się współzależnością. Im więcej pamięci, percepcji i prognozowania przechodzi do infrastruktury cyfrowej, tym mniej sensowne jest pytanie o to, czy „AI zastąpi człowieka", gdyż właściwym obiektem analizy staje się nowy system złożony.

Dopiero na tym tle można przejść do kwestii AGI bez ulegania mitologii jednego punktu osobliwości. Rzeczywiste wdrożenia pokazują bowiem, że kompetencja nie jest jednowymiarową własnością, autonomia nie jest synonimem inteligencji, a rezultat społeczny wynika z interakcji wielu podmiotów. Następnym krokiem musi być zatem krytyczna rekonstrukcja samego pojęcia Artificial General Intelligence: jego historii, definicji, kryteriów ogólności i benchmarków, a także relacji do autonomii oraz problemu „ruchomej bramki", która przesuwa się wraz z postępem technologii. Dopiero wtedy będzie można ocenić, czy AGI rzeczywiście stanowi właściwy horyzont badań, czy też koncepcja AGE ujawnia głębszy brak w sposobie, w jaki mierzymy rozwój sztucznej inteligencji.

> Być może największym paradoksem współczesnej technologii jest to, że nie musimy doczekać się narodzin świadomego bytu, by stać się od niego całkowicie zależnymi. Prawdziwa osobliwość nie nastąpi w momencie, gdy maszyna zacznie myśleć jak człowiek, lecz gdy człowiek przestanie potrafić funkcjonować bez maszynowego wsparcia poznawczego. Czy w świecie, gdzie percepcja i osąd zostają oddelegowane do infrastruktury, wciąż będziemy mogli mówić o autonomii podmiotu, czy staniemy się jedynie biologicznym ogniwem w cyfrowym łańcuchu sprawczym?

Inspiracje

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

Rana Gujral (ur. 18 czerwca 1976 r.) jest indyjsko-amerykańską przedsiębiorczynią, dyrektorem naczelnym i autorką specjalizującą się w sztucznej inteligencji, sztucznej inteligencji kognitywnej (AI) i obliczeniach afektywnych. Gujral pełnił funkcję dyrektora generalnego Behavioral Signals, firmy zajmującej się sztuczną inteligencją, która opracowuje oprogramowanie do analizy mowy i emocji oraz analizy zachowań, służące do komunikacji między ludźmi a maszynami. Przed pracą w Behavioral Signals, Gujral założył TiZE, firmę zajmującą się oprogramowaniem chmurowym dla przedsiębiorstw, która została przejęta przez Alchemy w 2016 roku. Pełnił również funkcje kierownicze i restrukturyzacyjne w firmach Cricut i Logitech. Jako lider i mówca w dziedzinie sztucznej inteligencji, Gujral zgłębia interakcję człowiek-maszyna, inteligencję głosową oraz koncepcję sztucznego doświadczenia ogólnego (AGE), badając, jak sztuczna inteligencja kształtuje intuicyjne podejmowanie decyzji, poznanie i przywództwo organizacyjne.

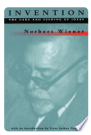
Rana Gujral *(born June 18, 1976) is an Indian-American entrepreneur, business executive, and author specializing in artificial intelligence, cognitive AI, and affective computing. Gujral served as the Chief Executive Officer of Behavioral Signals, an enterprise artificial intelligence company that develops speech-to-emotion and behavioral analytics software designed to bridge communication between humans and machines. Prior to Behavioral Signals, Gujral founded TiZE, an enterprise cloud software company that was acquired by Alchemy in 2016. He also held leadership and turnaround roles at Cricut and Logitech. As an AI thought leader and speaker, Gujral explores human-machine interaction, voice intelligence, and the concept of Artificial General Experience (AGE), examining how artificial intelligence shapes intuitive decision-making, cognition, and organizational leadership.*

Behavioral Signals

Literatura uzupełniająca

Książki

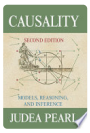
Książki autorów przywoływanych i literatura pokrewna:



[1] **"Invention: The Care and Feeding of Ideas"**

Norbert Wiener

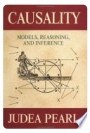
MIT Press | ISBN: 9780262731119



[2] "Causality"

Judea Pearl

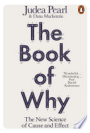
Cambridge University Press | ISBN: 9780521895606



[3] "Causality: Models, Reasoning and Inference"

Judea Pearl

Cambridge University Press | ISBN: 9780521773621



[4] "The Book of Why"

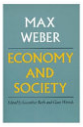
Judea Pearl

Penguin UK | ISBN: 9780241242643

[5] "The rational and social Foundations of music"

Max Weber

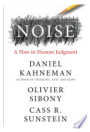
ISBN: 9780758136602



[6] "Economy and Society"

Max Weber

Univ of California Press | ISBN: 9780520035003



[7] "Noise: A Flaw in Human Judgment"

Daniel Kahneman

Hachette UK | ISBN: 9780316451383



[8] "The Adoption and Effect of Artificial Intelligence on Human Resources Management"

Pallavi Tyagi, Naveen Chilamkurti, Simon Grima, Kiran Sood, Balamurugan Baluswamy

2023 | Emerald Group Publishing | ISBN: 9781803820293



[9] "The Synthetic Mind Architecture"

Fouad Sabry

2026 | Thothora



[10] "Between the Species"

Arnold Arluke, Clinton Sanders

2009 | Allyn & Bacon



[11] "Why Machines Will Never Rule the World"

Jobst Landgrebe, Barry Smith

2022 | Taylor & Francis | ISBN: 9781000628692



[12] **"AI Vs Human Intelligence"**

Ashok Reddy Nemilidinne

2026 | Independently Published | ISBN: 9798198675483

Artykuły naukowe

Prace autorów przywoływanych w tekście:

[1] **"The Human Use of Human Beings: Cybernetics and Society"**

Stephen B. Miles, Norbert Wiener

1951 | Land Economics | DOI: 10.2307/3159747

Źródło

[2] **"The Human Use of Human Beings"**

Norbert Wiener

1950

Źródło

[3] **"IEEE 3rd Conference on Norbert Wiener in the 21st Century: Being Human in a Global Village"**

N. Wiener, G. Adamson

2020 | IEEE technology & society magazine | DOI: 10.1109/mts.2020.3002051

Źródło

[4] **"Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference"**

Judea Pearl

1988

Źródło

[5] **"Causality: Models, Reasoning and Inference"**

Christopher Hitchcock, Judea Pearl

2001 | The Philosophical Review | DOI: 10.2307/3182612

Źródło

[6] **"Causal Diagrams for Epidemiologic Research"**

Sander Greenland, Judea Pearl, James M. Robins

1999 | Epidemiology | DOI: 10.1097/00001648-199901000-00008

Źródło

[7] "Causal inference in statistics: An overview"

Judea Pearl

2009 | Statistics Surveys | DOI: 10.1214/09-ss057

Źródło

[8] "Cognition in the Wild"

Edwin L. Hutchins

1995 | The MIT Press eBooks | DOI: 10.7551/mitpress/1881.001.0001

Źródło

[9] "Distributed cognition in an airline cockpit"

Edwin L. Hutchins, Tove Klausen

1996 | Cambridge University Press eBooks | DOI: 10.1017/cbo9781139174077.002

Źródło

[10] "The social organization of distributed cognition."

Edwin L. Hutchins

1991 | American Psychological Association eBooks | DOI: 10.1037/10096-012

Źródło

[11] "Economy and Society: An Outline of Interpretive Sociology"

Max Weber

2002 | DOI: 10.1002/9780470755679.ch3

Źródło

[12] "FROM 'POLITICS AS A VOCATION'"

Max Weber

1991 | Edinburgh University Press eBooks | DOI: 10.1515/9781474469654-020

Źródło

[13] "Critical Note"

Arne M. Weber, Gottfried Vosgerau

2018 | The Oxford Handbook of 4E Cognition | DOI: 10.1093/oxfordhb/9780198735410.013.21

Źródło

Artykuły naukowe autora:

[14] "AI in the Real World: Applications and Use Cases"

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch03

Źródło

[15] "The AI Instinct"

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.fmatter

Źródło

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

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch02

Źródło

[17] "The Emergence of Artificial General Intelligence (AGI)"

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch04

Źródło

[18] "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: ab3699f0-f296-4964-9495-6d2241cebeb0