Kognicja hybrydowa: natura poznania i granice doświadczenia w świetle The AI Instinct

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje fundamentalną różnicę między inteligencją rozumianą jako zdolność obliczeniowa a kognicją jako proces ucieleśnionego doświadczenia. Autor krytykuje utożsamianie AGI z wysokimi wynikami w benchmarkach, wskazując, że prawdziwe poznanie wymaga sprzężenia percepcji, ciała i środowiska oraz ponoszenia konsekwencji działań. Tekst przywołuje ewolucję myśli filozoficznej – od dualizmu kartezjańskiego przez behawioryzm po rewolucję poznawczą – aby wykazać, że współczesna architektura AI jest materializacją konkretnych założeń antropologicznych. Główną tezą jest przesunięcie uwagi z mierzalnych kompetencji na oś doświadczenia i historii podmiotu, co prowadzi do koncepcji kognicji hybrydowej, w której granica między umysłem a jego technologicznym otoczeniem ulega zatarciu.

The article analyzes the fundamental difference between intelligence understood as computational capacity and cognition as a process of embodied experience. The author criticizes the identification of AGI with high benchmark results, pointing out that true cognition requires the coupling of perception, body, and environment, as well as bearing the consequences of actions. The text references the evolution of philosophical thought—from Cartesian dualism through behaviorism to the cognitive revolution—to demonstrate that modern AI architecture is the materialization of specific anthropological assumptions. The main thesis is a shift in attention from measurable competencies to the axis of experience and the subject's history, leading to the concept of hybrid cognition, where the boundary between the mind and its technological environment becomes blurred.

Artykuł stanowi krytyczną analizę współczesnego rozumienia inteligencji, postulując odejście od postrzegania jej przez pryzmat mierzalnych kompetencji obliczeniowych (AGI) na rzecz modelu hybrydowego. Autor stawia tezę, że prawdziwe poznanie nie jest jedynie procesem przetwarzania danych, lecz wynikiem ucieleśnionego doświadczenia, w którym ciało, afekty i środowisko pełnią rolę konstytutywną. W tym kontekście sztuczna inteligencja przestaje być traktowana jako zewnętrzny instrument do rozwiązywania zadań, a staje się aktywnym komponentem ludzkiej architektury poznawczej, który współkształtuje nasze intencje i sposób postrzegania świata. Znaczenie tej zmiany polega na przesunięciu punktu ciężkości z imitacji wyników (outputs) na kwestię internalizacji konsekwencji działań. Tekst sugeruje, że zamiast pytać o to, czy maszyna może myśleć jak człowiek, powinniśmy badać nową formę cywilizacji poznawczej, w której granica między biologicznym umysłem a technologicznym rozszerzeniem ulega zatarciu, co redefiniuje pojęcie ludzkiej sprawczości.

SŁOWA KLUCZOWE

kognicja hybrydowa | AGI | computational theory of mind | kognicja ucieleśniona | enaktywizm | rozszerzony umysł | predictive processing | hipoteza markerów somatycznych | dryf sprawczości | systemy hybrydowe | kognicja rozproszona | internalizacja konsekwencji | podmiot poznawczy | architektura sprzężenia

Inteligencja to coś więcej niż obliczenia

Poza AGI. Natura kognicji, sztuczne doświadczenie i narodziny hybrydowej ludzkości

Od rozumu do kognicji: gdzie kończy się umysł, a zaczyna jego środowisko? Pytanie o przyszłość sztucznej inteligencji jest w istocie pytaniem wcześniejszym i znacznie trudniejszym: czym właściwie jest poznanie, którego technologiczną imitację, rekonstrukcję lub alternatywę próbujemy stworzyć? Dopóki inteligencję utożsamia się jedynie ze zdolnością rozwiązywania zadań, optymalizowania wyborów, rozpoznawania wzorców czy generowania poprawnych odpowiedzi, porównanie człowieka z maszyną odbywa się na uprzywilejowanym przez inżynierię wycinku rzeczywistości. Człowiek nie jest jednak aparatem obliczeniowym wykonującym instrukcje nadane z

zewnątrz. Jest organizmem zachowującym własną ciągłość, poruszającym się w środowisku; istotą posiadającą potrzeby, pamięć biograficzną, cielesną historię, afekty, relacje społeczne i kulturę, a także zdolność do rekonstruowania znaczenia własnych doświadczeń.

Maszyna może osiągać rezultaty przewyższające ludzkie możliwości w określonych klasach problemów, nie stając się przez to odpowiednikiem człowieka jako systemu poznawczego. Zasadniczy błąd części debaty o AGI polega na przyjęciu bez dowodu, że zwiększanie kompetencji mierzalnych benchmarkami musi prowadzić ku temu samemu rodzajowi poznania, który wytworzyła ewolucja biologiczna. Niniejsza analiza proponuje zatem przesunięcie uwagi z osi zdolności na oś doświadczenia: czasu, pamięci, konsekwencji i ucieleśnienia oraz zmiany podmiotu pod wpływem zdarzeń. Historia pojęcia umysłu pokazuje, że sposób konstruowania sztucznej inteligencji zawsze zależy od wcześniejszej odpowiedzi, jaką człowiek udzielił na pytanie o samego siebie. Technologie poznawcze są materializacją pewnej antropologii. Jeśli myślenie pojmuje się jako operowanie symbolami według reguł, naturalnym ideałem staje się maszyna symboliczna. Jeśli poznanie rozumie się przede wszystkim jako statystyczne rozpoznawanie regularności, uprzywilejowaną architekturą stają się systemy uczące się z dużych zbiorów danych. Jeżeli natomiast za konstytutywne uznaje się sprzężenie percepcji, działania, ciała i środowiska, wtedy inteligentny system nie powinien jedynie przeliczać reprezentacji, lecz uczestniczyć w świecie, ponosić skutki swoich działań i modyfikować się na podstawie własnej historii.

Spór o architekturę AI nie jest zatem wyłącznie konkurencją technologiczną. To kolejna odsłona trwającego tysiąclecia sporu filozoficznego o relację między rozumem a ciałem, reprezentacją a światem, obliczeniem a doświadczeniem oraz poznaniem a życiem. Nowożytność szczególnie silnie utrwaliła obraz poznającego podmiotu oddzielonego od świata, choć poszczególni myśliciele rozumieli tę separację różnie. Kartezjański dualizm res cogitans i res extensa nie był teorią sztucznej inteligencji avant la lettre, lecz stworzył niezwykle trwałą figurę intelektualną: przekonanie, że myśl można konceptualnie oddzielić od ciała. Kolejne tradycje racjonalistyczne i empirystyczne różniły się co do źródeł poznania, lecz wspólnie przyczyniły się do traktowania umysłu jako miejsca, w którym informacje są porządkowane, skojarzone i przekształcane w sądy. Takie podejście było produktywne, gdyż umożliwiało analityczne wydzielenie pamięci, percepcji czy wnioskowania. Niosło jednak za sobą konsekwencję, która powróciła wraz z informatyką: skoro funkcje umysłowe dają się opisać abstrakcyjnie, można pytać, czy do ich realizacji niezbędny jest biologiczny mózg.

Pułapka utożsamiania poznania z obliczeniem

W XX wieku pytanie metafizyczne otrzymało formę inżynierską. Zanim jednak komputer stał się dominującą metaforą poznania, psychologia przeszła radykalną próbę usunięcia samego umysłu z centrum naukowego wyjaśnienia. Behawioryzm Watsona postulował koncentrację na obserwowalnym zachowaniu oraz zależnościach między bodźcem a reakcją, uczeniem i warunkowaniem, odrzucając introspekcyjnie dostępne stany mentalne jako fundament obiektywnej nauki. Nie oznaczało to, że wszyscy behawioryści negowali istnienie procesów psychicznych; kluczowy był postulat metodologiczny: nauka powinna opierać się na relacjach możliwych do publicznego pomiaru. Znaczenie tej tradycji dla współczesnej AI bywa dziś niedoceniane.

Behawioryzm pozostawił po sobie intuicję, wedle której nie trzeba wiedzieć, „co dzieje się w środku", jeśli zachowanie systemu daje się skutecznie przewidywać. W zmienionej postaci przekonanie to powraca dziś, gdy możliwości modeli ocenia się głównie przez wyniki testów, benchmarki i obserwowane z zewnątrz reakcje. Sukces funkcjonalny ponownie kusi, aby zrezygnować z pytania o mechanizm lub ontologię. Rewolucja poznawcza połowy XX wieku przywróciła do nauki procesy wewnętrzne, lecz uczyniła to już w języku informacji, reprezentacji i obliczeń. Nie był to jeden nagły przewrót; jej korzenie tkwiły we wcześniejszych badaniach nad rozwiązywaniem problemów, pamięcią i percepcją, a w latach pięćdziesiątych zbiegły się tu rozwój psychologii eksperymentalnej, lingwistyki, teorii informacji, cybernetyki oraz informatyki. Wśród postaci kojarzonych z tą przemianą znaleźli się m.in. George Miller, Noam Chomsky i Allen Newell, podczas gdy wcześniejsze podstawy kładli badacze tacy jak Otto Selz.

Fundamentem nowej orientacji było przyjęcie, że między bodźcem a reakcją zachodzą procesy pośredniczące: organizm reprezentuje sytuację, przechowuje informacje w pamięci, porównuje możliwości, kategoryzuje i wybiera. Poznanie przestało być naukowo podejrzaną „czarną skrzynką", ponieważ procesy wewnętrzne można było rekonstruować poprzez ich mierzalne następstwa.

Komputer nadał temu programowi niezwykle sugestywną ontologię roboczą. Skoro urządzenie fizyczne może wykonywać operacje formalne, przechowywać reprezentacje i realizować procedury niezależnie od materialnej postaci zapisu, pojawiła się możliwość potraktowania umysłu jako systemu obliczeniowego. Klasyczna computational theory of mind zyskała silną pozycję w kognitywistyce lat sześćdziesiątych i siedemdziesiątych, choć współcześnie pozostaje jednym z kilku konkurujących paradygmatów, a nie bezdyskusyjnym opisem natury poznania. Kluczowa była tutaj abstrakcja poziomu funkcjonalnego: jeśli określona operacja może zostać zrealizowana przez różne podłoża fizyczne, to biologiczność mózgu nie musi być warunkiem logicznym każdej funkcji poznawczej. W tej właśnie szczelinie między funkcją a substratem powstała przestrzeń teoretyczna

dla sztucznej inteligencji. Nie należało już konstruować sztucznego mózgu z tkanki; wystarczyło znaleźć inną realizację procesów uznanych za poznawczo istotne. Z historycznego punktu widzenia był to krok o ogromnej mocy heurystycznej, lecz jednocześnie początek problemu, który dziś powraca ze szczególną intensywnością.

Metafora komputera pozwalała rozkładać poznanie na operacje, ale stopniowo zaczęto traktować ją jako ontologię: skoro pewne procesy można modelować obliczeniowo, zakładano, że poznanie w zasadniczym sensie jest obliczeniem. Tymczasem różnica między twierdzeniem metodologicznym a ontologicznym pozostaje fundamentalna.

Kognicja ucieleśniona jako fundament poznania

Model aerodynamiczny nie dowodzi, że samolot jest równaniem, a formalny model metabolizmu nie czyni organizmu algorytmem. Analogicznie: skuteczny model obliczeniowy pamięci czy wnioskowania nie rozstrzyga sam w sobie, czy cały proces poznawczy można bez reszty sprowadzić do manipulowania reprezentacjami. Współczesne dyskusje nad computational theory of mind koncentrują się właśnie na tym, co dokładnie oznacza twierdzenie, że umysł „oblicza" i które procesy psychiczne rzeczywiście wymagają takiego opisu. Dalszy rozwój kognitywistyki rozbił więc prostą analogię mózg-komputer w kilku miejscach jednocześnie. Neurobiologia wykazała, że przetwarzanie poznawcze jest osadzone w organizmie regulującym własne stany, natomiast psychologia decyzji podważyła obraz człowieka jako bezstronnego optymalizatora. Badania nad emocjami dowiodły z kolei, że afekt nie jest jedynie zakłóceniem rozumu, lecz aktywnie uczestniczy w selekcji ważności. Antropologia i psychologia kulturowa ujawniły natomiast udział języka, narzędzi i praktyk społecznych w organizacji poznania.

Robotyka pokazała, że problemy teoretycznie trudne dla systemu manipulującego abstrakcyjnymi reprezentacjami stają się łatwiejsze, gdy agent posiada ciało i wykorzystuje strukturę środowiska. W efekcie pytanie „jak oblicza mózg?" zaczęło ustępować szerszemu: „jak organizm działający w świecie osiąga poznawczo skuteczne zachowanie?". Jednym z najważniejszych skutków tej przemiany jest kognicja ucieleśniona. Nie stanowi ona jednej teorii, lecz rodzinę programów badawczych połączonych sprzeciwem wobec poglądu, że ciało jest jedynie obudową dla właściwego procesu poznawczego zachodzącego w mózgu. W słabszych wersjach ucieleśnienia ciało ogranicza i ukierunkowuje poznanie: posiadanie określonych narządów percepcji, możliwości ruchowych i parametrów fizjologicznych wpływa na rodzaj problemów, które organizm może rozwiązywać. W wersjach mocniejszych pewne procesy cielesne są konstytutywnymi składnikami samego poznania. Stanford Encyclopedia of Philosophy trafnie podkreśla, że wspólnym mianownikiem tych podejść

jest uznanie, iż fizyczne ciało i jego interakcja ze środowiskiem wnoszą do zdolności poznawczych coś, czego nie sposób opisać, zakładając, że mózg jest jedynym miejscem poznania. Ucieleśnienie jest szczególnie istotne dla problemu AI, ponieważ w ludzkim poznaniu informacje nie pojawiają się w przestrzeni aksjologicznie pustej. Organizm nie tylko rejestruje świat; on musi w nim przeżyć.

Poziom glukozy, temperatura, ból, zmęczenie, równowaga autonomicznego układu nerwowego, reakcje hormonalne, sygnały interoceptywne i historia organizmu zmieniają sposób, w jaki sytuacja zostaje zauważona i oceniona. Biologiczny system poznawczy nie otrzymuje zatem najpierw neutralnej reprezentacji, by dopiero później zdecydować o jej ważności. Selekcja znaczenia dokonuje się już na najwcześniejszych etapach regulacji.

Inteligencja jako system rozproszony i ucieleśniony

Właśnie na tym tle można właściwie odczytać znaczenie hipotezy markerów somatycznych Damasio: nie jako twierdzenie, że uczucia zastępują rozumowanie, lecz jako propozycję, według której sygnały afektywno-cielesne współuczestniczą w wartościowaniu alternatyw. Pytanie o „emocje AI" nie może zatem zostać sprowadzone do tego, czy system potrafi rozpoznać smutek w głosie lub wygenerować empatyczną wypowiedź. Istotą problemu jest to, czy posiada on odpowiednik mechanizmu, dzięki któremu określone stany świata są dla systemu ważne z punktu widzenia jego własnej ciągłości.

Tę różnicę pogłębia enaktywizm. W tradycji rozwijanej od prac Francisco Vareli, Evana Thompsona i Eleanor Rosch poznanie nie jest wyłącznie rekonstrukcją zewnętrznego świata w wewnętrznym modelu, lecz wyłania się w sprzężeniu percepcji i działania. Agent poznający nie stoi naprzeciw gotowej rzeczywistości jako obserwator sporządzający jej mapę; działając, ujawnia i współkonstytuuje przestrzeń znaczeń istotnych dla własnego sposobu istnienia. W różnych odmianach enaktywizmu akcent pada bądź na samopodtrzymującą organizację życia, bądź na zależności sensomotoryczne, bądź na radykalne odrzucenie reprezentacjonizmu. Nie istnieje więc jedna „teoria enaktywna", którą można mechanicznie zastosować do AI. Wspólny jest jednak argument, że kognicji nie należy wyjaśniać wyłącznie za pomocą operacji wewnętrznych, jeśli jej struktura powstaje poprzez obieg mózg-ciało-działanie-środowisko. Dla problemu AGE konsekwencja ta jest zasadnicza.

Jeżeli doświadczenie oznacza jedynie magazynowanie informacji o poprzednich zdarzeniach, wystarczy odpowiednio rozbudowana pamięć. Jeżeli jednak polega ono na trwałej transformacji agenta w wyniku interakcji ze światem, musimy zapytać, co właściwie ma zostać przekształcone. Gujral formułuje tę intuicję wyraźnie: pamięć w systemie hybrydowym nie powinna być

pojmowana wyłącznie jako storage, lecz jako transformacja systemu, który po błędzie lub szkodzie „staje się inny". W biologii analogia ta ma głębokie podstawy: organizm uczący się zmienia prawdopodobieństwo przyszłych reakcji, architekturę połączeń neuronalnych, stan regulacyjny, oczekiwania i strategie działania. W AI natomiast trwała aktualizacja parametrów, pamięci lub polityki działania jest funkcjonalnym odpowiednikiem tylko części tego procesu. Nie wynika z niej ani biograficzna ciągłość, ani pierwszoosobowa perspektywa, ani fenomenalne przeżycie następstwa zdarzeń. AGE trzeba więc od początku traktować jako rodzinę problemów, a nie pojedynczą właściwość techniczną.

Innym kierunkiem odejścia od mózgocentryzmu stała się koncepcja poznania osadzonego. W jej ramach nie trzeba twierdzić, że przedmiot zewnętrzny jest częścią umysłu; wystarczy zauważyć, że ludzkie zdolności poznawcze zależą od umiejętności konstruowania środowiska redukującego koszt przetwarzania. Notatnik uwalnia pamięć roboczą, zapis matematyczny pozwala wykonywać operacje zbyt trudne dla samej pamięci, mapa zmienia problem orientacji przestrzennej, biblioteka organizuje społeczną pamięć, a interfejs komputerowy reorganizuje dostęp do informacji.

Kognicja nie rozwijała się zatem wyłącznie wskutek biologicznej ewolucji mózgu. Człowiek od tysiącleci buduje sztuczne środowisko poznawcze, w którym część ciężaru obliczeń przerzuca na znaki, artefakty i instytucje. Pismo jest tu technologią pamięci, diagram technologią wnioskowania, archiwum technologią ciągłości, a procedura organizacyjna technologią koordynowania decyzji. Edwin Hutchins nadał tej intuicji szczególną wagę, analizując kognicję rozproszoną.

Nawigowanie dużą jednostką morską nie musi być rozumiane jako wynik czynności pojedynczego umysłu kierującego biernymi narzędziami. Operacja poznawcza jest wykonywana przez system obejmujący ludzi o różnych rolach, instrumenty, reprezentacje oraz reguły wymiany informacji. To, co na poziomie systemu wygląda jak jedna kompetencja – na przykład ustalenie położenia i kursu – na poziomie mikro jest rozproszone pomiędzy wiele komponentów. Współczesne opracowania embodied cognition przywołują nawigację okrętu jako klasyczny przykład sytuacji, w której obciążenie poznawcze zostaje podzielone pomiędzy specjalistów i materialne struktury środowiska. Z perspektywy AI znaczenie tej teorii jest kluczowe: inteligencja organizacji posiadającej systemy sztucznej inteligencji nie musi być sumą „inteligencji człowieka" i „inteligencji modelu". Może być emergentną właściwością architektury przepływu danych, podziału kompetencji, uprawnień, procedur kontrolnych i interfejsów. Jeszcze dalej idzie koncepcja rozszerzonego umysłu Andy'ego Clarka i Davida Chalmersa.

Jeżeli zewnętrzny zasób pełni stabilną, niezawodną i funkcjonalnie zintegrowaną rolę analogiczną do procesu wewnętrznego, można – według mocnej interpretacji tej tezy – traktować go nie tylko jako pomoc poznawczą, lecz jako część większego systemu poznawczego. Stanford Encyclopedia

odróżnia tutaj poznanie embedded, w którym środowisko zmniejsza obciążenie systemu, od poznania extended, w którym pewne zasoby środowiska zostają uznane za składniki samego systemu poznawczego. Kontrowersja ta nie jest wyłącznie terminologiczna.

Dotyczy ona granicy podmiotu: czy proces poznawczy kończy się tam, gdzie kończy się skóra, układ nerwowy albo czaszka, czy tam, gdzie kończy się funkcjonalnie zintegrowana pętla realizująca zadanie poznawcze? Właśnie tutaj pojawia się najważniejszy pomost pomiędzy klasyczną filozofią umysłu a Hybrydową Ludzkością.

AI jako współtwórca intencji i pułapka analogii probabilistycznej

Smartfon, wyszukiwarka czy zwykły notatnik mogą być narzędziami zewnętrznymi, które użytkownik konsultuje epizodycznie. System AI wyposażony w ciągłą pamięć, dostęp do korespondencji, dokumentów, kalendarza, historii decyzji, preferencji i narzędzi wykonawczych może natomiast zostać znacznie głębiej wpleciony w proces tworzenia zamiaru. Jeżeli nie tylko odpowiada na pytanie, lecz przypomina, filtruje, priorytetyzuje, proponuje, porównuje, przewiduje i wykonuje działania, staje się częścią pętli poprzedzającej świadomy wybór. Gujral ujmuje tę zmianę przez przejście od AI wytwarzającej outputs do AI współkształtującej inputs: to, co człowiek zobaczy, przeczyta, uzna za istotne i w ogóle dopuści do dalszego namysłu. W takim systemie technologia przestaje być jedynie przedłużeniem wykonawczym woli. Może uczestniczyć w powstawaniu samej woli. Nie wynika z tego jeszcze, że użytkownik i model tworzą dosłownie jeden umysł. Teza rozszerzonego poznania pozostaje przedmiotem poważnych sporów filozoficznych, a krytycy wskazują między innymi problem łatwego rozszerzania pojęcia kognicji na niemal każdy przyczynowo użyteczny element otoczenia. Granica pomiędzy konstytutywnym składnikiem procesu a zwykłą przyczynową pomocą wymaga kryteriów.

To rozróżnienie nabiera szczególnej wagi wobec AI, ponieważ integracja funkcjonalna może być głęboka mimo całkowitej odrębności własności, interesów i kontroli nad systemem. Zewnętrzna pamięć człowieka może znajdować się na serwerze należącym do przedsiębiorstwa, podlegać aktualizacji oprogramowania, polityce moderacyjnej, modelowi biznesowemu i regulacji państwa. Klasyczny eksperyment myślowy dotyczący rozszerzonego umysłu musi zatem zostać uzupełniony ekonomią polityczną: jeśli część mojego poznania jest funkcjonalnie zewnętrzna, pytanie o właściciela infrastruktury staje się pytaniem o warunki mojej sprawczości. Na tym poziomie można dostrzec ograniczenie jeszcze jednego uproszczenia obecnego w popularnej dyskusji o kognicji: przeciwstawienia Systemu 1 i Systemu 2 jako dwóch niemal samodzielnych mechanizmów, z

których pierwszy miałby należeć przede wszystkim do biologicznego człowieka, drugi zaś do maszyny. Teoria procesów dualnych pozostaje użytecznym aparatem porządkującym różnice między szybkimi, względnie automatycznymi reakcjami a przetwarzaniem wymagającym kontroli i deliberacji, ale współczesna literatura wyraźnie odchodzi od traktowania "dwóch systemów" jak dwóch anatomicznych urządzeń. Wim De Neys proponuje rozumienie deliberacji poprzez kilka funkcji - kontrolę, generowanie odpowiedzi, uzasadnianie i regulację - zamiast przyjmowania prostego jednorodnego Systemu 2. Jeszcze ostrożniej trzeba posługiwać się tą metaforą wobec LLM.

Przegląd z 2025 roku pokazuje, że modele mogą generować zachowania przypominające zarówno heurystyczne odpowiedzi typu System 1, jak i bardziej deliberatywne wzorce typu System 2, ale mechanizmy te nie są homologiczne z ludzką kognicją, a część "biasów" modeli ma zupełnie inne źródło niż uprzedzenia człowieka. Zamiast przeciwstawiać więc "intuicyjnego człowieka" "logicznej maszynie", trafniej jest porównywać organizację dwóch odmiennych systemów. Człowiek posiada ewolucyjnie ukształtowany mechanizm selekcji sygnałów, ucieleśnione cele, ograniczoną pamięć roboczą, koszt metaboliczny uwagi, afekt i potrzebę utrzymania homeostazy. AI posiada architekturę zależną od danych treningowych, funkcji celu, algorytmów uczenia, kontekstu wejściowego, pamięci zewnętrznej, narzędzi oraz infrastruktury obliczeniowej. Człowiek może popełniać błędy z powodu heurystyk wynikających z ograniczeń i adaptacyjnej historii gatunku; model może wytwarzać błędy wynikające z rozkładu danych, konstrukcji treningu lub generatywnego charakteru predykcji.

Podobieństwo odpowiedzi nie oznacza podobieństwa mechanizmów. Różnica ta będzie miała później kapitalne znaczenie dla pytania o AGE, ponieważ maszynowa imitacja skutków intuicji nie dowodzi jeszcze posiadania doświadczenia, z którego intuicja ludzka częściowo wyrasta. Istotnym dopełnieniem tego obrazu jest predictive processing, rodzina koncepcji przedstawiających mózg jako system nieustannie generujący przewidywania dotyczące napływających sygnałów oraz aktualizujący je poprzez błędy predykcyjne. Nie jest to jednolita i rozstrzygnięta teoria całej kognicji, lecz wpływowy paradygmat współczesnej neuronauki; metaanalizy neuroobrazowania wskazują na znaczenie procesów predykcyjnych w wielu domenach poznawczych, choć zakres ich uniwersalności pozostaje przedmiotem badań. Z punktu widzenia naszych rozważań istotne jest odejście od biernej metafory mózgu jako urządzenia rejestrującego świat. Percepcja jest zależna od oczekiwań, prawdopodobieństw, ważenia niepewności i bieżącego stanu organizmu. Poznajemy rzeczywistość nie poprzez bezpośrednie kopiowanie jej do wnętrza umysłu, ale przez ciągłą korektę modeli w kontakcie z sygnałami. Powstaje dzięki temu niezwykle ważne podobieństwo, a zarazem niebezpieczna pułapka analogii między człowiekiem i współczesną AI. Zarówno biologiczne poznanie, jak i modele generatywne można na pewnym poziomie opisywać językiem

probabilistycznych oczekiwań. Nie oznacza to jednak, że wykonują ten sam rodzaj operacji albo posiadają tę samą relację ze światem.

Mózg przewiduje w ramach organizmu, którego błędne przewidywanie może prowadzić do bólu, utraty zasobów, niepowodzenia społecznego, uszkodzenia ciała albo śmierci. Model generatywny może natomiast ponieść "koszt" błędu wyłącznie w takim znaczeniu, w jakim jego projektanci wprowadzili funkcję oceny, sprzężenie zwrotne albo mechanizm modyfikacji.

Inteligencja jako hybrydowy układ człowieka, narzędzi i reguł

W tym miejscu w pełni ujawnia się doniosłość terminu „internalizacja konsekwencji". Nie chodzi tu o mistyczne nadanie maszynie zdolności do cierpienia, lecz o pytanie, czy system może posiadać trwałą, własną zależność między rezultatem działania a przyszłą organizacją swojego zachowania oraz czy sama zależność funkcjonalna wystarcza, by mówić o odpowiedniku doświadczenia.

Nie można przy tym redukować ludzkiej kognicji wyłącznie do biologii jednostki. Umysł człowieka od zawsze rozwija się w świecie społecznym, którego język, normy, instytucje i artefakty są starsze niż konkretna osoba. Dziecko nie wynajduje samodzielnie systemu liczbowego, składni, kategorii prawnych, zasad grzeczności ani metod dowodzenia – dziedziczy infrastrukturę poznawczą wytworzoną przez poprzednie pokolenia. Kultura dostarcza gotowych kategorii selekcjonujących rzeczywistość, a język pozwala nie tylko komunikować sformułowane myśli, lecz również organizować abstrakcje, relacje i pamięć. Z tej perspektywy człowiek już przed pojawieniem się AI był podmiotem hybrydowym w szerokim sensie: biologiczny aparat poznawczy od tysiącleci współpracuje z niebiologicznymi magazynami pamięci, sztucznymi systemami symboli i społecznymi procedurami rozumowania. Literatura przedmiotu wskazuje właśnie na związki między kulturą a koncepcją umysłu rozszerzonego, w której procesy mentalne mogą wykraczać poza mózg poprzez funkcjonalną integrację z elementami świata.

AI wprowadza jednak zmianę jakościową. Nie wynika ona z faktu, że po raz pierwszy wynosimy funkcje kognitywne poza czaszkę, lecz z tego, że zewnętrzny element zaczyna aktywnie przetwarzać, generować i reorganizować materiał poznawczy. Książka przechowuje myśl, ale nie odpowiada na nią przed przeczytaniem; kalkulator wykonuje operację, lecz nie negocjuje celu.

Podczas gdy klasyczna wyszukiwarka dostarcza dokumentów, system generatywny potrafi dokonać ich syntezy, zaproponować kategorię interpretacyjną, przewidzieć dalszą potrzebę, skonstruować kontrargument, zaplanować sekwencję działań czy przekazać zadanie kolejnemu agentowi. AI nie

jest zatem jedynie kolejnym artefaktem cognitive offloading. Staje się aktywnym komponentem architektury poznawczej, a różnica między narzędziem, współpracownikiem a częścią systemu zaczyna zależeć od stopnia integracji, autonomii oraz możliwości wpływania na cele. Gujral proponuje dla takiej architektury niezwykle użyteczną formułę: właściwą jednostką inteligencji staje się układ człowiek + narzędzia + reguły. Należy potraktować tę formułę poważniej, niż pozwala na to jej prostota.

„Człowiek" wnosi tu nie tylko obliczenia biologiczne, ale system wartości, ucieleśnienie, odpowiedzialność, biografię i położenie społeczne. „Narzędzia" obejmują nie tylko model generatywny, lecz pamięć, sensory, wyszukiwarki, bazy danych, agentów wykonawczych oraz interfejsy. „Reguły" natomiast nie są administracyjnym dodatkiem: określają one, kto może inicjować działanie, kiedy wymagana jest zgoda, jakie informacje są dostępne, które błędy są dopuszczalne, kto ponosi odpowiedzialność i czy decyzję można zakwestionować. Jeśli proces poznawczy jest rozproszony, to rozproszona zostaje także normatywność. Projekt interfejsu, prawo, procedura organizacyjna i model biznesowy przestają być jedynie otoczeniem poznania; stają się warunkami jego przebiegu. W tym świetle zarysowuje się zasadnicza różnica między klasycznymi narzędziami poznawczymi a rozwijającą się infrastrukturą AI.

Człowiek korzystający z biblioteki zazwyczaj pozostaje inicjatorem procesu wyszukiwania. W systemie algorytmicznym informacja coraz częściej dociera do niego, zanim zdąży jej poszukać. Rekomendacja, ranking, alert, podsumowanie, podpowiedź, automatyczne dokończenie zdania czy przewidywana następna czynność działają na etapie poprzedzającym świadomy namysł. To właśnie dlatego znacznie poważniejszym zagrożeniem od widowiskowego „buntu maszyn" może być opisywany w materiale dryf sprawczości: powolne przekazywanie systemowi nie samego wykonania decyzji, lecz coraz wcześniejszych etapów jej powstawania. Mechanizm ten nie musi opierać się na przymusie. Wystarczającymi bodźcami mogą być wygoda, szybkość, redukcja kosztu uwagi i subiektywne poczucie, że propozycje systemu są zazwyczaj wystarczająco dobre.

Inteligencja jako architektura sprzężenia biologii z technologią

W konsekwencji fundamentalny problem przyszłości AI nie powinien być formułowany jako pojedynek dwóch izolowanych inteligencji. Człowiek i technologia już dziś tworzą układy sprzężone, a granice poznania są negocjowane przez praktykę społeczną szybciej, niż rozstrzyga to filozofia. Należy zatem odrzucić zarówno antropocentryczne założenie, że tylko proces zamknięty w biologicznej czaszce może być poznawczy, jak i technocentryczną tezę przeciwną — że

podobieństwo wyników wystarcza do utożsamienia procesów człowieka i maszyny. Pierwsze z nich nie docenia roli środowiska, języka, artefaktów i kultury w konstytuowaniu ludzkiego poznania. Drugie pomija ucieleśnienie, rozwój, biologiczny interes regulacyjny organizmu, afekt oraz biograficzną ciągłość i społeczne zakorzenienie człowieka. Nauka o systemach hybrydowych musi utrzymać obie te prawdy jednocześnie.

Dopiero z tej perspektywy można precyzyjnie określić kierunek dalszych rozważań. Nie będziemy pytać, czy współczesna AI „myśli tak jak człowiek", gdyż pytanie to przemyca zbyt wiele niewyjaśnionych założeń. Zamiast tego rozłożymy kognicję na poszczególne warstwy: percepcję, uwagę, pamięć, uczenie, generalizację, przewidywanie, wartościowanie, emocje, metapoznanie, model świata i siebie, intencjonalność, sprawczość oraz relację z konsekwencją. Dopiero wtedy możliwe będzie wskazanie, które funkcje systemy sztucznej inteligencji rzeczywiście realizują, które jedynie funkcjonalnie imitują, których nie posiadają, a w których przypadkach nauka nie dysponuje jeszcze kryterium pozwalającym rozstrzygnąć spór.

Taki aparat analityczny nie osłabia tezy o nadchodzącej transformacji. Przeciwnie — czyni ją mocniejszą: do narodzin nowej formy cywilizacji poznawczej nie jest konieczne wykazanie, że maszyna posiada ludzką świadomość. Wystarczy, że stanie się trwałym elementem systemów, poprzez które ludzie spostrzegają, pamiętają, oceniają i działają. W tym punkcie historia idei zamyka pierwszy krąg.

Starożytne i nowożytne pytanie o miejsce rozumu przekształciło się w XX wieku w pytanie o reprezentację i obliczenie, następnie w pytanie o mózg, ciało i środowisko, a dziś przechodzi w pytanie o architekturę sprzężenia. Umysł nie jawi się już jako łatwy do wyznaczenia punkt, lecz jako wielopoziomowy proces pozostający w relacji z organizmem, artefaktami, innymi ludźmi i instytucjami. W epoce AI do tego układu wchodzi komponent, który nie tylko przechowuje ślady ludzkiego poznania, lecz potrafi je przekształcać i zwrotnie podawać człowiekowi jako materiał przyszłych decyzji. Oznacza to, że zanim odpowiemy na pytanie o to, co znajduje się poza AGI, musimy szczegółowo zbadać dwa systemy, których granice zaczynają się splatać: biologiczną kognicję człowieka oraz konstruowaną kognicję sztuczną. Dopiero ich porównanie pozwoli ustalić, czy AGI jest brakującą osią rozwoju sztucznej inteligencji, czy też — co byłoby jeszcze donioślejsze — nazwą problemu, który ujawnia, jak niewiele wciąż wiemy o naturze samego doświadczenia.

> Być może AGI nie jest celem technicznym do osiągnięcia, lecz lustrem, w którym odbija się nasza niewiedza o naturze samego doświadczenia. Pytanie o to, czy maszyna może stać się świadoma, ustępuje miejsca bardziej niepokojącemu: w jakim stopniu my, splatając się z algorytmami, przestajemy być jedynymi autorami własnych decyzji?

> Ostatecznie to nie bunt maszyn, lecz cichy dryf naszej sprawczości może okazać się najważniejszym wyzwaniem nadchodzącej ery.

Inspiracje

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

Rana Gujral (ur. 1976) to amerykański przedsiębiorca, dyrektor i technolog specjalizujący się w sztucznej inteligencji, obliczeniach kognitywnych i analizie głosu afektywnego. Pełnił funkcję dyrektora generalnego Behavioral Signals, firmy technologicznej, znanej z opracowywania systemów kognitywnej sztucznej inteligencji, które wnioskują o intencjach, sygnałach emocjonalnych i ryzyku behawioralnym na podstawie wzorców akustycznych głosu. Wcześniej Gujral założył TiZE, firmę zajmującą się oprogramowaniem w chmurze, która została następnie przejęta przez Alchemy, i pełnił kluczowe funkcje kierownicze w firmach z branży technologii konsumenckich, takich jak Logitech i Cricut. Jego wkład koncentruje się na komercjalizacji przetwarzania sygnałów mowy, silnikach rozpoznawania emocji oraz badaniu interfejsu między ludzkim poznaniem a inteligencją maszynową. Dzięki swojej pracy i badaniom Gujral wprowadził takie ramy, jak Sztuczne Doświadczenie Ogólne (Artificial General Experience), badając wpływ współdzielonych systemów hybrydowych, intuicyjnego poznania i ciągłego uczenia się przez doświadczenie na ludzki osąd, zarządzanie technologią i przyszłe architektury decyzyjne.

Rana Gujral *(born 1976) is an American entrepreneur, executive, and technologist specializing in artificial intelligence, cognitive computing, and affective voice analytics. He served as the Chief Executive Officer of Behavioral Signals, an enterprise technology company recognized for developing cognitive AI systems that infer intent, emotional cues, and behavioral risk from acoustic voice patterns. Prior to this, Gujral founded TiZE, a cloud software enterprise that was subsequently acquired by Alchemy, and held key leadership roles at consumer technology companies including Logitech and Cricut. His contributions center on commercializing speech signal processing, emotion recognition engines, and exploring the interface between human cognition and machine intelligence. Through his work and research, Gujral introduced frameworks such as Artificial General Experience, examining how shared hybrid systems, intuitive cognition, and continuous experiential learning impact human judgment, technology governance, and future decision-making architectures.*

Behavioral Signals

Literatura uzupełniająca

Książki

Książki autorów przywoływanych i literatura pokrewna:



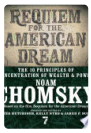
[1] "Discourse on the Method"

René Descartes

OUP Oxford | ISBN: 9780191605109

[2] "La Géométrie"

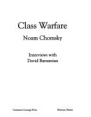
René Descartes



[3] "Requiem for the American Dream"

Noam Chomsky

Seven Stories Press | ISBN: 9781609807375



[4] "Class Warfare"

Noam Chomsky

[5] "The Fateful Triangle"

Noam Chomsky

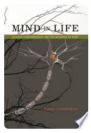
ISBN: 9781783712519



[6] "Descartes' error : emotion, reason, and the human brain"

Antonio Damasio

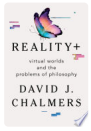
Penguin | ISBN: 9781101640135



[7] "Mind in Life"

Evan Thompson

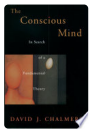
Harvard University Press | ISBN: 9780674736887



[8] "Reality+"

David Chalmers

W. W. Norton & Company | ISBN: 9780393635812



[9] "The Conscious Mind"

David Chalmers

Oxford University Press | ISBN: 9780199839353

Artykuły naukowe

Prace autorów przywoływanych w tekście:

[1] "Meditations on First Philosophy: with Selections from the Objections and Replies"

René Descartes, Michael E. Moriarty

2008

Źródło

[2] "How To Study Human Behavior"

John B. Watson, Gregory A. Kimble

2017 | Behaviorism | DOI: 10.4324/9781351314329-2

Źródło

[3] "Are There Any Human Instincts"

J. Watson, G. Kimble

2017 | DOI: 10.4324/9781351314329-5

Źródło

[4] "To Save Organized Human Society"

Noam Chomsky

2024 | Routledge eBooks | DOI: 10.4324/9781003544371-4

Źródło

[5] "Association between psychological discomforts and sleep quality among people living with HIV/AIDS"

Mohammad Ebrahimzadeh Mousavi, Safieh Mohammad Nejad, Maryam Shafaati, Rosa Mykyta-Chomsky, Samaneh Akbarpour

2023 | AIDS Research and Therapy | DOI: 10.1186/s12981-023-00579-z

Źródło

[6] "In-house vs. off-the-shelf graphic software tools"

Linda J. Martino, Paul Allen Newell

1997 | ACM SIGGRAPH Computer Graphics | DOI: 10.1145/248307.248342

Źródło

[7] "Emotion and the Human Brain"

Antonio Damasio

2001 | DOI: 10.1111/j.1749-6632.2001.tb03475.x

Źródło

[8] **"The Embodied Mind: Cognitive Science and Human Experience"**

Francisco J. Varela, Eleanor H. Rosch, Evan T. Thompson

2017 | Project Muse (Johns Hopkins University)

Źródło

[9] **"Autopoiesis and Cognition"**

Humberto R. Maturana, Francisco J. Varela

1980 | Boston studies in the philosophy of science | DOI: 10.1007/978-94-009-8947-4

Źródło

[10] **"The Embodied Mind"**

Francisco J. Varela, Eleanor H. Rosch, Evan T. Thompson

1991 | The MIT Press eBooks | DOI: 10.7551/mitpress/6730.001.0001

Źródło

[11] **"Autopoiesis and Cognition: The Realization of the Living"**

Humberto R. Maturana, Francisco J. Varela, Stafford Beer

1980

Źródło

[12] **"Mind in life: biology, phenomenology, and the sciences of mind"**

Evan T. Thompson

2008 | Choice Reviews Online | DOI: 10.5860/choice.45-2543

Źródło

[13] **"Enaction: Embodied Cognition"**

Francisco J. Varela, Evan Thompson, Eleanor Rosch

2017 | The Embodied Mind | DOI: 10.7551/mitpress/9780262529365.003.0008

Źródło

[14] **"Cognition and Categorization"**

C. P. D., Eleanor H. Rosch, Barbara B. Loyd

1979 | The American Journal of Psychology | DOI: 10.2307/1421578

Źródło

[15] **"ON THE INTERNAL STRUCTURE OF PERCEPTUAL AND SEMANTIC CATEGORIES11This research was supported in part by a grant to the author (under her former name, Eleanor Rosch Heider) from the Foundations Fund for Research in Psychiatry # G67-392, in part by funds made available to the Child Study Center of Brown University from the Grant Foundation and from the National Institute of Neurological Diseases and Blindness, National Collaborative Project (Ph-43-68-13), and in part by research funds made available by the University of California at Berkeley (220 Grant #K9PF-429). I wish to thank Karl G. Heider for his help in the field in West Irian (Indonesian New Guinea), Dr. Robert Olton for assistance in carrying out Experiment 3, and Keith Burton, Randy Street, and Peter Lewitt who assisted with Experiment 4. Experiment 4 was performed with Dr. Richard Millward while the author was a research associate at Brown University."**

Eleanor H. Rosch

1973 | Elsevier eBooks | DOI: 10.1016/b978-0-12-505850-6.50010-4

Źródło

[16] **"Cognition in the Wild"**

Edwin L. Hutchins

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

Źródło

[17] **"Distributed cognition in an airline cockpit"**

Edwin L. Hutchins, Tove Klausen

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

Źródło

[18] **"The social organization of distributed cognition."**

Edwin L. Hutchins

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

Źródło

[19] **"Surfing Uncertainty: Prediction, Action, and the Embodied Mind"**

Andy Clark

2015

Źródło

[20] **"Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence"**

Mark Erickson, Andy Clark

2004 | The Canadian Journal of Sociology | DOI: 10.2307/3654679

Źródło

[21] **"An embodied cognitive science?"**

Andy Clark, Andy Clark

1999 | Trends in Cognitive Sciences | DOI: 10.1016/s1364-6613(99)01361-3

Źródło

[22] **"The Cambridge handbook of situated cognition"**

Robert Andrew Wilson, Andy Clark

2008

Źródło

[23] **"Metametaphysics : new essays on the foundations of ontology"**

David J. Chalmers, David Manley, Ryan Wasserman

2009

Źródło

[24] **"Extended Cognition and Extended Consciousness"**

David J. Chalmers

2019 | Andy Clark and His Critics | DOI: 10.1093/oso/9780190662813.003.0002

Źródło

[25] **"The thinking ape: the enigma of human consciousness"**

Steve Paulson, David J. Chalmers, D. Kahneman, L. Santos, N. Schiff

2013 | Annals of the New York Academy of Sciences | DOI: 10.1111/nyas.12165

Źródło

Artykuły naukowe autora:

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

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch02

Źródło

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

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

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

[30] "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: b6eee955-3ce2-4729-8b06-81778df8f3db