Architektura Supersociety: od inteligencji jednostkowej do cywilizacji wielopodłożowej w kontekście The AI Instinct

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Tekst analizuje przejście od indywidualnej kognicji do struktur Supersociety, wprowadzając rozróżnienie między podmiotowością fenomenalną a funkcjonalną. Autor omawia koncepcję Multi-Substrate Person oraz powstanie agentowych instytucji, które dzięki AI przekształcają pasywną biurokrację w aktywny system kognitywny. Kluczowym problemem jest ryzyko instytucjonalnej hipertrofii poznawczej i drastycznej asymetrii informacji między systemem a obywatelem. W odpowiedzi proponowane jest wprowadzenie 'constitutional collectiva intelligence' – ram prawnych i etycznych, które mają zapobiec autorytaryzmowi technologicznemu. Celem jest stworzenie hybrydowej inteligencji zbiorowej, która zwiększa efektywność społeczną bez niszczenia autonomii jednostki, pluralizmu oraz fundamentalnych praw obywatelskich w świecie zintegrowanych systemów AI.

The text analyzes the transition from individual cognition to Supersociety structures, introducing a distinction between phenomenal and functional subjectivity. The author discusses the concept of the Multi-Substrate Person and the emergence of agentic institutions that, through AI, transform passive bureaucracy into an active cognitive system. A key issue is the risk of institutional cognitive hypertrophy and drastic information asymmetry between the system and the citizen. In response, the introduction of 'constitutional collective intelligence' is proposed—legal and ethical frameworks intended to prevent technological authoritarianism. The goal is to create a hybrid collective intelligence that increases social efficiency without destroying individual autonomy, pluralism, or fundamental civil rights in a world of integrated AI systems.

Artykuł rzuca wyzwanie popularnym wizjom superinteligencji, odrzucając scenariusz narodzin pojedynczego, wszechmocnego bytu AI lub metafizycznej fuzji świadomości. Autor stawia tezę, że przyszłość kognicji nie leży w ontologicznej zmianie natury umysłu, lecz w ewolucji ludzkości ku hybrydowemu systemowi poznawczemu. W tej perspektywie AI przestaje być zewnętrznym narzędziem, a staje się aktywną warstwą infrastruktury cywilizacyjnej – rodzajem „rozproszonego Metamind", który integruje pamięć i sprawczość instytucji, państw i ludzi. Kluczowym przesunięciem jest tu przejście od technicznego problemu 'alignmentu' (dostrojenia modelu) do poziomu politycznego i konstytucyjnego zarządzania zbiorową inteligencją. Tekst argumentuje, że największym wyzwaniem nie będzie zatem walka z autonomiczną maszyną, lecz zaprojektowanie takiej architektury kognitywnej, która zwiększy możliwości cywilizacji, nie niszcząc przy tym pluralizmu, wolności jednostki i demokratycznej legitymizacji władzy. To zaproszenie do myślenia o superinteligencji nie jako o nowym gatunku, lecz jako o nowej skali organizacji ludzkiej wiedzy i sprawczości.

SŁOWA KLUCZOWE

Supersociety | cywilizacja wielopodłożowa | The AI Instinct | podmiotowość funkcjonalna | Multi-Substrate Person | Metamind | collective intelligence | hybrid collective cognition | Explanation Engine | constitutional collective intelligence | civilizational cognition | cognitive circuit breakers | governance of disagreement | alignment hybrydowy

Podmiotowość funkcjonalna zamiast metafizyki świadomości

Dopiero w pełni ukształtowany Metamind oznaczałby jedną podmiotowość fenomenalną i osobową realizowaną przez wiele substratów. Współczesna nauka nie daje jednak podstaw do rozstrzygnięcia tej kwestii; nie istnieje empiryczny test, który pozwoliłby stwierdzić, że dostatecznie zintegrowana sieć ludzi i maszyn tworzy jeden strumień świadomości. Wymagałoby to najpierw rozwiązania problemów omówionych w części VII, a następnie wykazania, że warunki świadomości mogą być realizowane na poziomie systemu rozproszonego.

Analogicznie należy stopniować pojęcie Multi-Substrate Person. W ujęciu prawnym może ono oznaczać człowieka korzystającego z komponentów technicznych uznawanych za przedłużenie jego funkcji. W wersji funkcjonalnej – osobę, której pamięć i działanie są trwale współrealizowane przez zewnętrzną infrastrukturę. Z kolei w wersji fenomenalnej mówimy o jednym świadomym podmiocie rozciągniętym pomiędzy substratami. Dziś jedynie dwie pierwsze wersje można sensownie rozwijać jako przedmiot badań i prawa, co ma kluczowe znaczenie dla kwestii odpowiedzialności. Jeśli osoba korzysta z protezy poznawczej, odpowiedzialność może nadal spoczywać na człowieku i dostawcach infrastruktury, zgodnie z wcześniej zdefiniowanymi zasadami. Gdybyśmy jednak mieli do czynienia z rzeczywistym podmiotem wielosubstratowym, klasyczne prawo stanęłoby przed problemem jego osobowości prawnej, ciągłości, własności i śmierci. Źródłowa wizja słusznie dostrzega, że utrata jednego substratu mogłaby wówczas zmienić sens tożsamości i odjścia z tego świata. Są to jednak pytania filozoficzne, przygotowujące grunt pod przyszłe spory, a nie opis aktualnego statusu technologicznego. Organizacje już dziś są osobami prawnymi, choć nie uznajemy ich za osoby fenomenalne – spółka może posiadać majątek, zobowiązania i zachować ciągłość niezależnie od wymiany pracowników.

Prawo zna zatem kategorię podmiotowości funkcjonalnej bez konieczności zakładania świadomości. Przyszłe autonomiczne organizacje hybrydowe mogą wymagać rozwoju właśnie tej tradycji, zamiast gwałtownego przechodzenia do metafizyki sztucznej osoby. Prowadzi to do pojęcia agentowej instytucji. Wyobraźmy sobie organizację, w której ludzie ustalają normy i cele strategiczne, lecz większość bieżącej alokacji zadań, monitorowania ryzyka, dokumentowania, prognozowania i negocjacji przejmują systemy sztuczne. Taka instytucja posiada trwałą pamięć, działa w trybie ciągłym, monitoruje środowisko i autonomicznie eskaluje problemy. System ten jest jakościowo bardziej „poznawczy” niż klasyczna biurokracja, mimo że nie musimy przypisywać mu świadomości. Max Weber opisywał biurokrację jako technologię racjonalizacji opartą na procedurach, hierarchii, specjalizacji i dokumentacji. Agentowa AI może stworzyć biurokrację kognitywnie aktywną, w której dokumentacja przestaje być pasywnym archiwum, a staje się elementem aktywnie przywoływanym przy każdej decyzji. Reguła nie tylko widnieje w podręczniku; agent może monitorować jej przestrzeganie w czasie rzeczywistym. Instytucja zaczyna pamiętać i reagować z dużo większą ciągłością niż jej ludzcy uczestnicy. Może to zwiększać sprawność państwa, firmy czy organizacji naukowej, ale rodzi również problem instytucjonalnej hipertrofii poznawczej. Organizacja dysponująca niemal doskonałą pamięcią i zdolnością predykcji może uzyskać przytłaczającą przewagę nad pojedynczym człowiekiem.

Konieczność konstytucyjnych ram dla zbiorowej inteligencji

Asymetria informacji między instytucją a obywatelem może w przyszłości ulec jeszcze drastyczniejszemu pogłębieniu. Supersociety musi zatem stać się przedmiotem analizy politycznej, gdyż nie każda wyższa inteligencja zbiorowa przekłada się na wzrost wolności. Autorytarne państwo może przecież funkcjonować jako doskonale skoordynowany system informacyjny; totalna integracja danych często zwiększa efektywność, jednocześnie niszcząc autonomię jednostki. Inteligencja systemu i dobro jego członków to odrębne zmienne.

Jeśli mierzymy Supersociety wyłącznie zdolnością do przewidywania epidemii, alokowania zasobów czy rozwiązywania problemów, za najbardziej inteligentny uznamy system wysoce scentralizowany. Z perspektywy politycznej jego jakość zależy jednak od praw członków, możliwości sprzeciwu, prywatności, pluralizmu oraz rozdziału władzy. Hybrydowa Ludzkość potrzebuje zatem pojęcia constitutional collective intelligence. Wysoka zdolność kognitywna społeczeństwa powinna być osadzona w ramach konstytucyjnych ograniczeń, które określą, jakich informacji system nie może wykorzystywać, jakie decyzje wymagają osobistego mandatu, które mniejszościowe stanowiska muszą zachować prawo do ekspresji i które warstwy pamięci nie mogą zostać scentralizowane.

Analogia do biologii organizmu jest tu zasadna tylko w bardzo ograniczonym zakresie. Komórka nie posiada prawa sprzeciwu wobec organizmu – obywatel tak. Dlatego metafora superorganizmu, stosowana normatywnie do społeczeństwa, może być niebezpieczna. Demokratyczne społeczeństwo nie powinno stać się organizmem w sensie całkowitego podporządkowania jednostek jednemu interesowi całości. Można zatem stwierdzić, że optymalne Supersociety powinno być poznawczo zintegrowane, ale politycznie niezredukowane do jednego podmiotu. Wspólna inteligencja ma służyć wielu osobom, a nie zamieniać je w funkcjonalne komórki jednego superorganizmu.

Tutaj ujawnia się głęboka różnica między collective intelligence a collective agency. Grupa może wspólnie rozwiązać problem, nie posiadając jednej woli. Collective agency pojawia się dopiero wtedy, gdy istnieje mechanizm reprezentujący wspólny cel i zdolność działania w imieniu całości – taką strukturę posiadają państwa i korporacje. Supersociety oparte na AI mogłoby znacznie zwiększyć szybkość i zakres tych mechanizmów.

Wraz z tym pojawia się pytanie o legitymizację: kto upoważnił system do działania w imieniu zbiorowości? W polityce demokratycznej legitymizacja wynika z instytucji przedstawicielskich, prawa i procedur. Algorytmiczna trafność nie może ich zastąpić. System przewidujący preferencje

społeczne lepiej niż sondaż nadal nie posiada demokratycznego mandatu, podobnie jak Metamind znający obywatela lepiej niż urzędnik nie uzyskuje automatycznie prawa do decydowania o jego dobru. Jest to szczególnie istotne w kontekście wizji superinteligencji jako fuzji, którą Gujral przeciwstawia obrazowi zewnętrznej ASI. Oficjalny opis książki przedstawia hybrydową kognicję i „superintelligence as fusion" jako jedne z trzech głównych idei autora.

Wizja ta eliminuje klasyczny problem „my versus machine", ale nie rozwiązuje kwestii władzy. Jeśli superinteligencja jest systemem człowiek-AI, wciąż musimy pytać, którzy ludzie i które AI – oraz według jakich reguł – tworzą tę hybrydę. Nie istnieje bowiem „człowiek" jako jednolity partner dla AI. Ludzkość jest pluralistyczna, nierówna ekonomicznie i politycznie, podzielona na państwa, klasy, organizacje i kultury.

Hybrydowa superinteligencja dostępna jedynie niewielkiej elicie może pogłębić dystans między ludźmi bardziej niż dystans między człowiekiem a maszyną. Sam Gujral w publicznych wypowiedziach wskazuje na ryzyko koncentracji dostępu do rozszerzeń kognitywnych i możliwość powstania podziału między osobami augmentowanymi a nieaugmentowanymi. Z tego powodu pojęcie Supersociety powinno obejmować dystrybucję mocy kognitywnej. Nie wystarczy znać średnią zdolność systemu; trzeba wiedzieć, kto posiada dostęp do kluczowych rozszerzeń, kto kontroluje ich konfigurację i kto może używać zbiorowej inteligencji przeciwko innym. Można to wyrazić poprzez rozróżnienie między collective capability a capability distribution.

Społeczeństwo jako całość może odnotować ogromny wzrost zdolności diagnostycznych czy naukowych, jednocześnie coraz bardziej nierówno rozdzielając możliwość korzystania z tych zasobów. Z punktu widzenia agregatów system staje się „inteligentniejszy", lecz z perspektywy wielu obywateli ich względna sprawczość maleje. W takim scenariuszu Hybrydowa Ludzkość nie tworzy jednego Metamind, lecz hierarchię sprzężeń. Najbogatsi dysponują prywatnymi systemami wieloagentowymi, organizacje własnymi modelami i pamięcią, państwa infrastrukturą strategiczną, a pojedynczy obywatel ma jedynie ograniczony dostęp do warstwy konsumenckiej. Jest to scenariusz znacznie bardziej realistyczny niż jednolite „zlanie się" ludzkości z AI.

Nauka o sieciach wskazuje ponadto, że w systemach kolektywnych pozycja w strukturze może być równie ważna jak kompetencja elementu. Węzeł kontrolujący przepływ informacji między częściami systemu uzyskuje ogromny wpływ, nie będąckoniecznie najlepszym problem solverem. W hybrid collective cognition pojawiają się analogiczne role – na przykład ludzcy gatekeeperzy oddzielający lub łączący podsieci agentów AI. Władza przyszłego Metamind może zatem spoczywać nie tam, gdzie występuje największa inteligencja lokalna, lecz tam, gdzie znajduje się kontrola interfejsów.

Superinteligencja jako policentryczny system polityczno-konstytucyjny

Wszystko to sprowadza się do ekonomii politycznej rozszerzonego umysłu. Dostawca API, system pamięci, protokół komunikacji czy meta-agent routujący zadania mogą mieć większe znaczenie niż pojedynczy model. Architektura Supersociety jest w istocie architekturą rynku i władzy. W tym ujęciu Explanation Engine należy traktować jako dobro publiczne. Jeśli jego celem jest organizowanie wiedzy całego społeczeństwa, nie może być on prywatnym silnikiem ustanawiającym domyślną teorię świata. Bardziej odporną formą byłby ekosystem wielu silników wyjaśniających, instytucji naukowych, publicznych repozytoriów, modeli i ludzi, których wyniki można wzajemnie konfrontować. Explanation Engine powinien przypominać konstytucyjnie zorganizowaną wspólnotę naukową, a nie wyrocznię; jego „moc wyjaśniająca" wynikałaby z zachowania błędu, sporu i procedury krytyki wewnątrz systemu.

Prowadzi to do paradoksu: najbardziej inteligentny Metamind może być tym, który potrafi nie być jednym umysłem. Jeśli jedność oznaczałaby całkowitą homogenizację, system straciłby różnorodność niezbędną do korekty. Jeśli natomiast Metamind zostanie zdefiniowany jako sieć utrzymująca spójny model działania, lecz zachowująca niezależne centra wiedzy i krytyki, będzie mógł łączyć szybkość koordynacji z odpornością epistemiczną. Takie podejście współgra z koncepcją policentryzmu. Inteligencja nie musi być centralnym mózgiem; może być zdolnością federacji do koordynowania wielu autonomicznych centrów. Właśnie dlatego Supersociety bardziej przypominałoby internet niż pojedynczy komputer, choć internet sam w sobie nie posiada jednej funkcji celu. Możliwe byłoby zbudowanie kolejnych warstw koordynacji, które zwiększałyby wspólną sprawczość bez całkowitego centralizowania informacji.

Warto odróżnić skalę kognicji od jej szybkości. System może integrować ogromne ilości informacji, reagując jednak wolno z powodu proceduralnych zabezpieczeń. Inny może działać niemal natychmiastowo, lecz w oparciu o uproszczony model. W zarządzaniu kryzysowym istnieje stałe napięcie pomiędzy deliberacją a szybkością, dlatego Supersociety musiałoby posiadać różne tryby działania, dopasowane do rodzaju problemu. Można wyobrazić sobie system pracujący hierarchicznie: rutynowe problemy rozwiązywane są lokalnie, sygnały nietypowe eskalują do szerszych struktur, konflikt wartości uruchamia procedurę polityczną, a niepewność poznawcza kieruje sprawę do sieci ekspertów. Takie hierarchiczne zarządzanie niepewnością byłoby bardziej realistyczne niż jeden uniwersalny superagent. Tutaj z miejsca zastosowanie znajdują cognitive circuit breakers. Na poziomie społecznym ich rolę pełniłyby instytucje hamujące automatyczną konwergencję: sądy, audyty, niezależna nauka, komisje, wolne media, procedury odwoławcze i

konkurujące modele. Supersociety pozbawione takich bezpieczników mogłoby być bardzo szybkie i bardzo niebezpieczne. Wizja superinteligencji hybrydowej jako systemu zdolnego przekraczać biologiczne ograniczenia człowieka i unikać zagrożenia zewnętrznej ASI jest optymistyczna, lecz wymaga istotnego zastrzeżenia: fuzja nie usuwa problemu alignment, tylko przenosi go do wnętrza hybrydy.

Konieczne staje się uzgodnienie celów nie tylko AI z człowiekiem, ale różnych ludzi między sobą, jednostki z organizacją, interesów krótkoterminowych z długoterminowymi oraz lokalnych preferencji z prawami wspólnoty. Dlatego alignment na poziomie Metamind przypomina bardziej politykę konstytucyjną niż dostrojenie pojedynczego modelu. Społeczeństwa od tysiącleci rozwiązują problem sprzecznych celów nie poprzez tworzenie jednego idealnego systemu wartości, lecz za pomocą norm, prawa, procedur kompromisu i instytucji ograniczających władzę. Hybrydowa superinteligencja również będzie potrzebowała governance of disagreement. Może to być kluczowa różnica pomiędzy systemem sztucznym a społeczeństwem. Pojedynczego agenta można teoretycznie zoptymalizować według jednej funkcji celu, podczas gdy społeczeństwo składa się z wielu częściowo sprzecznych koncepcji dobra.

Próba redukcji tej wielości do jednej funkcji może być politycznie groźniejsza niż niedoskonałość samej AI. Dlatego Supersociety nie powinno być projektowane jako „globalny optymalizator", lecz jako narzędzie zwiększające zdolność ludzi do rozwiązywania sporów przy zachowaniu wielości celów. Jest to mniej spektakularne niż jednolity superumysł, ale bardziej zgodne z naturą społeczeństw. Najwyższą formą kolektywnej inteligencji może być zatem nie zdolność do natychmiastowego znalezienia jednej odpowiedzi, lecz umiejętność rozpoznania, które pytania w ogóle nie mają jednej odpowiedzi technicznej. Niektóre problemy są optymalizacyjne, inne normatywne, a jeszcze inne wymagają politycznego kompromisu pomiędzy dobrami niewspółmiernymi. Explanation Engine musi umieć odróżniać te kategorie. W przeciwnym razie nastąpi technokratyczna redukcja problemów wartości do problemów przewidywania. System może doskonale oszacować konsekwencje trzech wariantów polityki, ale nie rozstrzygnie samym faktem trafniejszej predykcji, jaki rozkład kosztów jest sprawiedliwy.

Superinteligencja jako hybrydowa kognicja cywilizacyjna

Metamind może wspierać deliberację, jednak jego rola nie powinna sprowadzać normatywnej decyzji do pozornie neutralnego outputu. Na tym tle pojawia się fascynująca możliwość: zbiorowa metakognicja cywilizacyjna. Społeczeństwo mogłoby dysponować systemami monitorującymi

jakość własnej wiedzy, które wykrywałyby obszary głębokiej niepewności, identyfikowały rozbieżności między modelami, mapowały zależności infrastrukturalne i wskazywały domeny, w których zbyt silnie polegamy na jednym źródle. Byłaby to forma świadomości własnych ograniczeń, bez konieczności zakładania istnienia jednej, fenomenalnej świadomości całej ludzkości. Tak rozumiany Metamind mógłby komunikować funkcjonalnie: „w tym obszarze nasza wiedza jest słaba", „te dwie instytucje opierają się na tym samym źródle danych", „ten model stanowi pojedynczy punkt awarii" lub „ta hipoteza ma niewystarczające potwierdzenie". Nie jest to introspekcja cywilizacji w sensie metafizycznym, lecz metakognitywna infrastruktura zarządzania wiedzą. To właśnie tutaj AI może wnieść zdolność dotychczas niedostępną ludzkości.

Żaden człowiek nie jest w stanie przeglądać całej globalnej literatury i śledzić wszystkich eksperymentów, jednocześnie analizując relacje między nimi. Systemy sztuczne mogą potencjalnie podtrzymywać mapę wiedzy obejmującą znacznie większą część pamięci cywilizacyjnej. Jednocześnie taka infrastruktura stwarza bezprecedensowe ryzyko centralizacji epistemicznej. Jeśli jedna organizacja zyska kontrolę nad globalną mapą wiedzy, problem staje się polityczny. Dlatego najbardziej wartościowa metakognicja powinna być federacyjna, audytowalna i wielomodelowa. Wtedy można wrócić do najśmielszej tezy źródła: że inteligencja może w przyszłości przekraczać ograniczenia biologicznego substratu i kontynuować się w układach wielopodłożowych.

Ten proces już trwa. Wiedza człowieka od tysięcy lat przeżywa swoich autorów dzięki językowi, pismu, bibliotekom, instytucjom i technologii. AI nie rozpoczyna zatem „ucieczki wiedzy z biologii", lecz radykalnie zwiększa aktywność niebiologicznej części pamięci cywilizacyjnej. Książka przechowuje wiedzę biernie; model może ją przeszukiwać, łączyć i wykorzystywać do działania. Pamięć instytucjonalna w archiwach wymaga człowieka, który ją odnajdzie; agent może stale zestawiać bieżącą decyzję z tysiącami wcześniejszych przypadków. Ta aktywizacja zewnętrznej pamięci jest znacznie bardziej empirycznie uchwytnym sensem przejścia ku „wielopodłożowej inteligencji" niż hipotetyczne przenoszenie osobowej świadomości. Można zatem mówić o civilizational cognition: ludzkość od dawna poznaje za pomocą sieci ludzi, narzędzi i instytucji, jednak AI gwałtownie zwiększa szybkość, automatyzm i samodzielność niebiologicznych elementów tej sieci. „Coś poza AGI" może zatem powstać bez jednego, konkretnego momentu narodzin superinteligentnej maszyny. Najpierw pojawią się prawdopodobnie organizacje, laboratoria i państwa funkcjonujące jako coraz bardziej zintegrowane systemy hybrydowe, których zbiorowa zdolność znacznie przewyższy możliwości każdego pojedynczego człowieka czy modelu.

Dopiero później pojawi się pytanie, czy integracja osiągnęła poziom uzasadniający nowe kategorie ontologiczne. Prowadzi to do istotnego odwrócenia klasycznej narracji singularity. Zamiast jednego punktu zwrotnego, po którym pojedyncza AI staje się „mądrzejsza od ludzkości", możemy

zaobserwować rozproszoną zmianę skali kognicji, w której sama ludzkość staje się coraz bardziej hybrydowym systemem obliczeniowo-instytucjonalnym.

Granica epoki nie musi więc przypominać narodzin nowego podmiotu, lecz stopniową przebudowę wszystkich istniejących podmiotów i instytucji. W takim scenariuszu pojęcie superinteligencji również wymaga rewizji. Jeśli inteligencja jest własnością systemu, pytanie „czy AI przewyższy człowieka?" jest błędnie postawione na poziomie analizy. Właściwsze brzmi: które konfiguracje ludzi, modeli, agentów, pamięci i instytucji będą posiadały najwyższą zdolność rozwiązywania problemów i jak zostanie rozdzielona władza nad tymi konfiguracjami? To pytanie przesuwa superinteligencję z obszaru ontologii do socjologii i ekonomii politycznej.

Supersociety jako ekologia kognitywna i rozproszona inteligencja

Najbardziej kompetentnym „bytem" może okazać się nie robot, lecz organizacja. Najbardziej wpływową formą inteligencji – infrastruktura łącząca miliony lokalnych agentów. W takim ujęciu kluczowy problem alignmentu przestaje być relacją człowiek-maszyna, a staje się kwestią relacji człowieka z hybrydową organizacją i społeczeństwem. Źródłowa opozycja między „zewnętrzną ASI" a „hybrydowym człowiekiem" okazuje się zbyt binarna; istnieje przestrzeń dla dziesiątek stanów pośrednich, w tym dla wysoce autonomicznych systemów sztucznych funkcjonujących wewnątrz hybrydowych struktur.

Człowiek może być silnie rozszerzony przez jedne narzędzia i jednocześnie podporządkowany innym. Nie ma konieczności wyboru jednej trajektorii dla całej cywilizacji, dlatego bardziej użyteczne jest myślenie o ekologii kognitywnej niż o pojedynczym, finalnym superumyśle. Ekologia z natury obejmuje różne typy organizmów i relacji: konkurencję, symbiozę, pasożytnictwo, kooperację oraz nisze. Podobnie przyszłe środowisko inteligencji będzie składać się z ludzi, prywatnych systemów agentowych, organizacji, modeli państwowych, AI open-source, robotów, autonomicznych usług i zespołów hybrydowych.

Ich relacje nie będą jednolite, co nadaje tej metaforze przewagę normatywną. Ekosystem nie musi być maksymalnie zunifikowany, aby pozostać produktywnym; stabilność może wynikać właśnie z różnorodności i wzajemnych ograniczeń. Hybrydowa Ludzkość nie potrzebuje zatem jednego globalnego Fused Mind, lecz może opierać się na wielu częściowo połączonych Metaminds, które współpracują, konkurują i kontrolują się nawzajem. W tym punkcie pojęcie „Supersociety" można zrekonstruować w formie naukowo ostrożniejszej. Supersociety byłoby społeczeństwem, w którym

coraz większa część zbiorowych funkcji poznawczych – pamięć, prognozowanie, alokacja uwagi, modelowanie konsekwencji, koordynacja i metakontrola – jest współrealizowana przez ludzi i systemy sztuczne, bez konieczności uznawania całej zbiorowości za jeden świadomy podmiot. Taka definicja oddziela problem funkcjonalny od metafizyki.

Pozwala to badać konkretne wskaźniki: szybkość reakcji, jakość decyzji, odporność na błąd wspólny, różnorodność, możliwość korekty, efektywność pamięci, zdolność uczenia się i dystrybucję sprawczości. Można empirycznie porównywać społeczeństwa i organizacje bez rozstrzygania, czy posiadają one „umysł" w sensie fenomenalnym. Analogicznie Explanation Engine można zdefiniować jako wielopoziomowy system generowania i testowania modeli przyczynowych, łączący sztuczne przeszukiwanie przestrzeni hipotez z ludzką i instytucjonalną zdolnością do krytyki, eksperymentu oraz normatywnej interpretacji. Metamind może oznaczać trwały system integrujący wiele wyspecjalizowanych źródeł poznania przy zachowaniu wspólnej pamięci i metakontroli. Multi-Substrate Person powinien natomiast pozostać terminem zarezerwowanym dla znacznie mocniejszej, obecnie spekulacyjnej hipotezy o osobowej ciągłości przekraczającej pojedynczy biologiczny substrat.

Tak uporządkowane pojęcia przestają być jedynie futurystycznym słownikiem i stają się hierarchią hipotez o rosnącej sile ontologicznej. Collective intelligence jest zjawiskiem empirycznym. Hybrid collective intelligence to rozwijający się przedmiot badań. Metamind w słabym, funkcjonalnym sensie stanowi rozsądny model systemowy, Supersociety jest scenariuszem organizacyjnym, natomiast Multi-Substrate Person w mocnym sensie pozostaje hipotezą filozoficzną. To uporządkowanie jest kluczowe przed finałem artykułu.

Cały wywód rozpoczął się od pytania o naturę kognicji ludzkiej i sztucznej. Wykazano w nim, że sama kompetencja nie wyczerpuje problemu doświadczenia, a AGI nie wyczerpuje problemu przyszłości inteligencji. Okazuje się również, że jednostka inteligencji nie musi pozostawać stała. Być może największym błędem współczesnej debaty jest założenie, że przyszła inteligencja nadal będzie posiadać strukturę znaną z indywidualnego człowieka: jednego agenta, jedną pamięć, jeden zestaw celów i jedną granicę. Natura i społeczeństwo dostarczają wielu przykładów funkcjonalnych systemów o rozproszonej zdolności rozwiązywania problemów. Nie oznacza to, że „ludzkość stanie się jednym umysłem". Znacznie ostrożniejsze i zarazem bardziej interesujące twierdzenie brzmi: ludzkość może stać się cywilizacją, której najważniejsze akty poznawcze będą powstawały na poziomie sieci biologicznych i sztucznych elementów, a pojedynczy człowiek nie będzie już właściwą jednostką wyjaśniającą ich pełną zdolność.

Superinteligencja jako architektura zbiorowej samokontroli

Taki proces ma swoje precedensy. Żaden człowiek nie „wie” całej współczesnej fizyki; dyscyplina ta posiada tę wiedzę w sposób rozproszony – poprzez ludzi, literaturę, aparaturę, dane i instytucje. Podobnie żaden pracownik nie reprezentuje pełnej wiedzy globalnego przedsiębiorstwa; organizacja przechowuje ją dzięki podziałowi kompetencji i pamięci. AI może jednak sprawić, że ta rozproszona wiedza stanie się znacznie bardziej aktywna, skoordynowana i dostępna w czasie rzeczywistym. To właśnie może stanowić prawdziwe przejście jakościowe: od społeczeństwa posiadającego rozproszoną wiedzę do takiego, które potrafi ją dynamicznie aktywować. Metamind nie musiałby zatem rodzić się jako nowa osoba; mógłby powstawać jako nowa zdolność cywilizacji do używania samej siebie jako systemu poznawczego.

Wówczas jednak jeszcze większego znaczenia nabiera ostrzeżenie przed dryfem. Im bardziej funkcjonalna staje się zbiorowa hybryda, tym mniej widoczne mogą być pojedyncze decyzje projektowe określające jej strukturę. Społeczeństwo może przyzwyczaić się do domyślnego routingu, pamięci, agentów i filtrów tak, jak dziś przyzwyczajamy się do sieci energetycznej. Kto projektuje architekturę zbiorowego poznania, projektuje w istocie przyszłą zdolność społeczeństwa do rozumienia samego siebie. Dlatego nie wystarczy zapytać, czy Metamind będzie inteligentny. Należy pytać, czy będzie pluralistyczny, odwracalny, audytowalny, federacyjny i politycznie legitymizowany. Nie wystarczy pytać, czy Supersociety trafnie przewiduje; trzeba pytać, czy zachowuje prawa ludzi, których te przewidywania dotyczą. Nie wystarczy, aby Explanation Engine tworzył sprawne modele – musi pozostawiać przestrzeń dla modelu konkurencyjnego i eksperymentu falsyfikującego. Najważniejszym kryterium przyszłej superinteligencji może zatem okazać się nie jej maksymalna moc obliczeniowa, lecz zdolność do organizowania tej mocy bez niszczenia różnorodności i sprawczości komponentów, z których powstała.

To jest również punkt, w którym pojęcie mądrości odzyskuje znaczenie. Inteligencja pozwala znajdować rozwiązania, wiedza dostarcza reprezentacji świata, a metakognicja umożliwia ocenę ograniczeń własnej wiedzy. Mądrość obejmuje jednak coś więcej: pytanie o cele, konsekwencje i proporcje pomiędzy dobrami. System może być superinteligentny w sensie instrumentalnym, a mimo to nie posiadać struktury instytucjonalnej umożliwiającej mądre działanie. Hybrydowa Ludzkość będzie więc wymagała nie tylko większej mocy poznawczej, lecz architektury zbiorowej samokontroli. Najcenniejszym elementem Metamind może być system potrafiący powiedzieć całej sieci: „konwergujemy zbyt szybko”, „brakuje nam niezależnych danych”, „ten cel jest w konflikcie z prawami mniejszości” lub „nie wiemy wystarczająco dużo, aby działać”. Dopiero w takim ujęciu

można mówić o Explanation Engine jako o czymś więcej niż superwyszukiwarce czy modelu predykcyjnym. Jego najwyższą funkcją nie byłoby udzielanie odpowiedzi na każde pytanie, lecz organizowanie procesu, dzięki któremu społeczeństwo potrafi odróżnić to, co wie, od tego, co jedynie przewiduje, wierzy, zakłada lub wartościuje.

Przejście od inteligencji jednostkowej do cywilizacji wielopodłożowej

W tym miejscu wracamy do AGE. Doświadczenie społeczne nie musi być fenomenalnym „przeżyciem". Może być historią, która realnie modyfikuje przyszłą architekturę działania. Kryzys finansowy zmienia regulacje, katastrofa lotnicza procedury bezpieczeństwa, epidemia systemy sanitarne, a przełom naukowy programy badań. AI może sprawić, że taka instytucjonalna rekonsolidacja będzie szybsza i bardziej systematyczna. Supersociety posiadałoby wówczas Artificially Augmented Collective Experience, choć nie byłoby to AGE w sensie świadomego doświadczenia maszyny.

Być może właśnie ten poziom okaże się donioślejszy niż pytanie o to, czy pojedynczy model coś doświadcza. Rzeczywistość społeczna może zostać przekształcona przez systemy pozbawione qualiów, ale pozwalające instytucjom pamiętać, modelować i reagować na skalę dotąd nieosiągalną. Ostatecznie najbardziej produktywna interpretacja wizji Metamind nie polega więc na ogłoszeniu narodzin „nowego gatunku świadomości". Polega na dostrzeżeniu, że ewolucja kognicji może zmieniać poziom organizacji, na którym pojawiają się funkcjonalnie najpotężniejsze procesy poznawcze. Najpierw pojedyncze komórki tworzyły organizmy wielokomórkowe; organizmy społeczne budowały grupy; ludzie stworzyli instytucje, naukę i globalne sieci informacji. AI może pogłębić integrację tej ostatniej warstwy do tego stopnia, że nowe zdolności wyłonią się na poziomie całej sieci.

Analogia ewolucyjna nie jest dowodem na to, że nastąpi kolejny „major transition in individuality". Jest jednak użytecznym pytaniem badawczym: czy połączenie pamięci, agentowości, komunikacji i mechanizmów kontroli może wytworzyć na poziomie hybrydowych instytucji nowe formy funkcjonalnej indywidualności? Odpowiedź wymagałaby badań nad granicami systemu, jego zdolnością do samopodtrzymania, realizacją celów, ciągłością pamięci i adaptacją. Na obecnym etapie można sformułować odpowiedź ostrożną.

Tak, nowe jednostki funkcjonalne są możliwe i częściowo już powstają w zespołach oraz organizacjach hybrydowych. Nie wiemy jednak, czy kiedykolwiek staną się one jednym

fenomenalnym podmiotem. Nauka może badać pierwsze zjawisko, nie udając, że rozstrzygnęła drugie. Ta asymetria między tym, co mierzalne funkcjonalnie, a tym, czego nie potrafimy rozstrzygnąć fenomenologicznie, jest jednym z przewodnich motywów całego artykułu. Podobieństwo zachowania nie dowodzi identyczności ontologicznej. Kompetencja nie jest świadomością. Pamięć nie jest doświadczeniem. Integracja nie jest osobą. Inteligencja zbiorowa nie jest wspólnym „ja". A jednak to właśnie integracja funkcjonalna może zmienić świat szybciej i głębiej niż rozwiązanie tych problemów filozoficznych. Państwa, przedsiębiorstwa, nauka i ludzie będą działać poprzez systemy hybrydowe niezależnie od tego, czy filozofia świadomości osiągnie konsensus.

Dlatego pytanie artykułu nie powinno brzmieć: „co jest kolejnym akronimem po AGI?". Nie ma konieczności, by po AGI następowało AGE, ASI, MSP czy jakikolwiek inny pojedynczy szczebel technologicznej drabiny. Prawdziwa transformacja może nie być liniową sekwencją nowych rodzajów maszyn, lecz zmianą skali podmiotu poznającego.

Najpierw AI była narzędziem jednostki. Następnie stała się partnerem zespołu, potem infrastrukturą organizacji. Kolejnym etapem może być warstwa poznawcza cywilizacji – sieć systemów sztucznych i biologicznych, która nie posiada jednego centrum, lecz współuczestniczy w tym, co społeczeństwa pamiętają, przewidują, wyjaśniają i wybierają. Wtedy pytanie o przyszłość inteligencji należy odwrócić. Kluczowe nie będzie to, jak inteligentne staną się maszyny, ale jakim systemem poznawczym stanie się ludzkość, gdy AI wejdzie w trwałą architekturę jej pamięci, uwagi, nauki, gospodarki i instytucji. Na tym poziomie Fundacja Dobre Państwo musi odpowiedzieć na pytanie: co rzeczywiście znajduje się poza AGI? Odpowiedź nie może być kolejnym technologicznym hasłem.

Musi ona połączyć kognicję ludzką i sztuczną, AGE, świadomość, systemy hybrydowe, politykę, ekonomię, epistemologię i antropologię w jedną teorię przejścia od inteligencji jednostkowej do cywilizacji wielopodłożowej – a następnie wskazać, jakie miejsce pozostaje w niej dla człowieka jako istoty zdolnej nie tylko przetwarzać informacje, lecz również nadawać znaczenie temu, kim chce się stać.

> W świecie zdominowanym przez hybrydową kognicję, największym ryzykiem nie jest błąd w kodzie AI, lecz utrata zdolności do nadawania znaczeń. Gdy granica między jednostką a systemem ulegnie rozmyciu, pozostanie tylko jedno fundamentalne pytanie: czy w architekturze wielopodłożowej cywilizacji znajdzie się jeszcze miejsce dla człowieka jako istoty, która nie tylko przetwarza dane, ale świadomie decyduje o tym, kim chce się stać?

Inspiracje

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

Rana Gujral (ur. 1976) jest indyjsko-amerykańskim dyrektorem ds. technologii, przedsiębiorcą i inwestorem, specjalizującym się w sztucznej inteligencji, obliczeniach kognitywnych i analizie afektywnej głosu. Pełnił funkcję dyrektora generalnego Behavioral Signals, firmy zajmującej się sztuczną inteligencją, znanej z rozwoju technologii rozpoznawania emocji i analizy behawioralnej opartej na głosie. Wcześniej w swojej karierze Gujral założył firmę oprogramowania chmurowego TiZE (przejętą przez Alchemy) i piastował stanowiska kierownicze w działach produktów i inżynierii w dużych firmach technologicznych, takich jak Logitech, Cricut i Kronos. Jako częsty prelegent na międzynarodowych forach technologicznych, w tym TEDx i World Government Summit, w swojej pracy bada powiązania między poznaniem maszynowym a ludzkim procesem decyzyjnym, proponując takie ramy, jak sztuczne doświadczenie ogólne (AGE) i poznanie hybrydowe.

Rana Gujral *(born 1976) is an Indian-American technology executive, entrepreneur, and investor specializing in artificial intelligence, cognitive computing, and affective voice analytics. He served as the Chief Executive Officer of Behavioral Signals, an enterprise artificial intelligence company known for developing voice-based emotion recognition and behavioral analysis technology. Earlier in his career, Gujral founded the cloud software company TiZE (acquired by Alchemy) and held product and engineering leadership roles at major technology companies, including Logitech, Cricut, and Kronos. A frequent speaker at international technology forums, including TEDx and the World Government Summit, his work explores the intersection of machine cognition and human decision-making, proposing frameworks such as Artificial General Experience (AGE) and hybrid cognition.*

Behavioral Signals

Literatura uzupełniająca

Artykuły naukowe

Autorzy przywoływani:

[1] **"FROM ‘POLITICS AS A VOCATION’"**

Max Weber

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

Google Scholar

Artykuły pokrewne (Google Scholar):

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

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch02

Google Scholar

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

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch03

Google Scholar

[4] **"The AI Instinct"**

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.fmatter

Google Scholar

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

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch04

Google Scholar

[6] **"Introduction: Our Definition of “There” Is Wrong"**

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch01

Google Scholar

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: 154f0d3a-df8e-4551-95da-5f427e267319