Hybrydowa Ludzkość: Poza AGI ku architekturze wspólnego poznania w świetle The AI Instinct

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Tekst analizuje ewolucję roli człowieka w dobie AI, wskazując na przesunięcie z zadań wykonawczych ku meta poziomowi kontroli i projektowania architektury epistemicznej. Autor ostrzega przed ryzykiem pogłębienia hierarchii społecznych oraz manipulacji kognitywnej, postulując wprowadzenie nowej alfabetyzacji obywatelskiej i ochrony wolności poznawczej (cognitive liberty). Kluczowym elementem jest postulat symetrii dostępu do narzędzi AI w relacji obywatel-państwo. Artykuł proponuje zasadę proporcjonalności integracyjnej: im większa zależność człowieka od technologii, tym silniejsze powinny być jego prawa do kontroli i odwołania systemu. Wizja Hybrydowej Ludzkości nie zakłada fuzji świadomości, lecz wysoki stopień funkcjonalnego współuczestnictwa, gdzie technologia rozszerza możliwości ludzkie bez odbierania podmiotowości.

The text analyzes the evolution of the human role in the era of AI, pointing to a shift from executive tasks toward a meta-level of control and the design of epistemic architecture. The author warns against the risk of deepening social hierarchies and cognitive manipulation, advocating for the introduction of a new civic literacy and the protection of cognitive liberty. A key element is the demand for symmetry of access to AI tools in the citizen-state relationship. The article proposes the principle of integrative proportionality: the greater a human's dependence on technology, the stronger their rights to control and override the system should be. The vision of Hybrid Humanity does not assume a fusion of consciousness, but rather a high degree of functional co-participation, where technology extends human capabilities without stripping away agency.

Artykuł stanowi krytyczną analizę przejścia od postrzegania sztucznej inteligencji jako narzędzia do traktowania jej jako trwałej infrastruktury poznawczej, co autor definiuje mianem Hybrydowej Ludzkości. Główną tezą tekstu jest twierdzenie, że kluczowym wyzwaniem ery po AGI nie będzie sam przełom techniczny, lecz konieczność zaprojektowania nowej architektury cywilizacyjnej. W tej wizji wzrost mocy obliczeniowej i kognitywnej systemów hybrydowych musi zostać zrównoważony przez rozwój instytucjonalnej odpowiedzialności oraz rygorystyczną ochronę wolności poznawczej człowieka. Autor argumentuje, że bez świadomego zarządzania relacją między biologią a technologią – w tym nad prawem do metapoziomu kontroli i odwracalności integracji – ryzykujemy zastąpienie ludzkiej sprawczości przez niewidoczną dominację algorytmiczną. Tekst przekształca zatem pytanie o moment nadejścia superinteligencji w program badawczy dotyczący tego, jak zachować pluralizm i autonomię w świecie, gdzie granica między umysłem a maszyną staje się porowata.

SŁOWA KLUCZOWE

Hybrydowa Ludzkość | architektura wspólnego poznania | poza AGI | metapoziom kontroli | cognitive liberty | wolność poznawcza | zasada proporcjonalności integracyjnej | infrastruktura kognitywna | pluralizm poznawczy | fuzja funkcjonalna | agentowość | cognitive offloading | etyka odwracalności | interdyscyplinarny governance

Przejście od operacji wykonawczych do metapoziomu kontroli

Wiedza naukowa wymaga szczególnie wysokich standardów. Badacz może korzystać z AI do syntezy danych i generowania hipotez, lecz musi zachować zdolność weryfikacji dowodów, metod oraz wniosków. Jeśli proces powstawania artykułu stanie się zbyt złożony, by jakikolwiek człowiek mógł go w pełni objąć rozumem, nauka będzie zmuszona wypracować nowe instytucje odpowiedzialności rozproszonej. Możliwe jest, że w przyszłości część badań przejmą hybrydowe laboratoria, w których agenci autonomicznie proponują eksperymenty, aparatura je wykonuje, a modele analizują dane – podczas gdy ludzie pełnić będą funkcję metanadzoru i interpretacji. Nie

musi to oznaczać końca naukowca; może raczej zmienić jego rolę z wykonawcy wielu operacji na projektanta epistemicznej architektury.

Podobne przesunięcie ról może dotknąć wielu zawodów. Człowiek przechodzi od operacji pierwszego rzędu ku operacjom metapoziomu: definiowaniu celu, wyborowi kryteriów, kontroli uprawnień i zarządzaniu wyjątkami. Nie jest jednak pewne, że każdy pracownik stanie się „dyrygentem agentów". Organizacje mogą centralizować metapoziom, pozostawiając części kadr bardziej ograniczone role wykonawcze. Dowodzi to, że technologia nie determinuje automatycznie społecznego modelu pracy. Z tej perspektywy walka o przyszłość zatrudnienia może stać się w istocie walką o prawo do metapoziomu: kto ustala cele i parametry agentów? Kto ma dostęp do pamięci i analizy? Kto posiada prawo zakwestionować model?

Jeśli te kompetencje zostaną zarezerwowane dla wąskiej kadry, AI pogłębi hierarchie. Jeśli jednak zostaną szeroko rozproszone, wzrośnie sprawczość pracowników. Analogicznie rzecz ma się z demokracją.

AI może obniżyć koszt analizy projektów prawnych, budżetów i danych publicznych, zwiększając zdolność obywateli do kontroli władzy. Państwo jednak będzie korzystać z tych samych narzędzi. Wynik tego starcia zależy od symetrii dostępu do zdolności poznawczych. Hybrydowa Ludzkość może zatem wymagać nowej infrastruktury obywatelskiej AI – systemów reprezentujących interesy jednostki wobec administracji, pomagających rozumieć prawo i dane publiczne. Taka „demokratyzacja doradztwa" mogłaby stać się jednym z najbardziej egalitarnych skutków technologii.

Jednocześnie narasta problem manipulacji. System znający obywatela w sposób intymny może dostosowywać polityczną perswazję z niespotykaną dotąd precyzją. Ochrona demokracji będzie wymagała zatem nie tylko fakt-checkingu, ale norm dotyczących mikrotargetowanego wpływu kognitywnego i jawności interesów stojących za agentem. W tym punkcie neuroetyczne pojęcie cognitive liberty wykracza poza ramy neurotechnologii. Wolność poznawcza nie może ograniczać się do ochrony mózgu przed bezpośrednim odczytem; powinna obejmować warunki, w których człowiek formuje przekonania bez niewidocznej, spersonalizowanej ingerencji realizującej cudzy interes. Nie chodzi tu o prawo do życia całkowicie wolnego od wpływów – byłoby to niemożliwe.

Różnica tkwi w przejrzystości, asymetrii i kontroli. Rozmówca wpływa na nas, lecz jest widzialnym podmiotem z własnym stanowiskiem. System natomiast może ukrywać ekonomiczny cel za interfejsem osobistego doradcy. Z tego powodu Hybrydowa Ludzkość będzie potrzebowała nowej alfabetyzacji obywatelskiej: umiejętności pytania nie tylko o to, czy odpowiedź jest prawdziwa, ale dlaczego została mi pokazana właśnie ta wersja, kto kontroluje mojego pośrednika i jaki ma interes

w mojej decyzji. To poziom, na którym epistemologia i polityka stają się nierozdzielne. Kto organizuje informację, wpływa na władzę. Kto kontroluje pamięć, kształtuje tożsamość. Kto zarządza agentem, steruje działaniem. Własność infrastruktury poznawczej staje się zatem w coraz większym stopniu własnością warunków sprawczości.

Ostatecznym celem regulacji nie powinno być jednak utrudnianie integracji. Hybrydy mogą dać ludziom możliwości dziś niedostępne, szczególnie w medycynie, edukacji i nauce. Zadaniem prawa jest raczej dopilnowanie, aby wzrost funkcjonalnej zależności był równoważony wzrostem praw kontrolnych. Można tu sformułować zasadę proporcjonalności integracyjnej: im bardziej technologia staje się konstytutywna dla zdolności człowieka do funkcjonowania, tym silniejsze powinny być jego prawa do ciągłości, przenośności, kontroli i odwołania. Zwykła aplikacja rozrywkowa wymaga słabszych gwarancji niż system podtrzymujący komunikację osoby sparaliżowanej czy agent zarządzający osobistym archiwum. W ten sposób antropologia technologii prowadzi do prawa. Człowiek rozszerzony nie jest mniejszym podmiotem prawa; przeciwnie – im bardziej zależy od infrastruktury, tym silniej system prawny musi chronić warunki, dzięki którym zależność nie przerodzi się w dominację.

Na tym tle źródłowa wizja Fused Mind może zostać zinterpretowana w sposób naukowo dojrzalszy. Nie musi ona oznaczać jednej świadomości rozciągniętej między neuronem a serwerem, lecz wysoki stopień funkcjonalnego współuczestnictwa sztucznej infrastruktury w działaniu człowieka – stan, w którym pytanie o granice odpowiedzialności i sprawczości staje się ważniejsze niż fizyczne położenie komponentu.

Tak rozumiana fuzja jest już częściowo możliwa bez BCI, podczas gdy silniejsza forma fuzji fenomenalnej pozostaje hipotezą. Rozdzielenie to pozwala zachować twórczy potencjał książki bez naruszania rygoru naukowego. Analogicznie AGE można traktować jako heurystykę badawczą doświadczeniopodobnej historyczności, a nie dowód sztucznej subiektywności. Supersociety zaś może pozostać teorią zwiększonej zbiorowej kognicji, a nie zapowiedzią jednego superorganizmu.

Program badawczy cywilizacji hybrydowej zamiast prognoz AGI

Explanation Engine można zachować jako program współtworzenia i testowania wyjaśnień, a nie traktować go jako bezpośredni dostęp maszyny do metafizycznej prawdy. To właśnie dzięki takim korektom najbardziej śmiałe intuicje okazują się intelektualnie wartościowe — nie dlatego, że można je już uznać za fakty, lecz ponieważ wskazują luki w dominującej mapie pojęciowej. AGI nie

mówi przecież wszystkiego o pamięci, doświadczeniu, relacjach i czasie. Benchmarki nie wyczerpują tematu instytucji. Intelligence nie tożsame jest z mądrością. Automation nie definiuje sprawczości. AI safety natomiast nie rozstrzyga kwestii politycznej legitymizacji.

W ten sposób odpowiedź na pytanie „co poza AGI?" przestaje być pojedynczą prognozą, a staje się programem badawczym cywilizacji hybrydowej. Jego przedmiotem jest przejście od modelu maszyny jako narzędzia do modelu sztucznej infrastruktury jako trwałego uczestnika ludzkiej kognicji, relacji i instytucji. Kluczowym postulatem tego programu byłoby ustalenie, jak zachować ludzką zdolność do refleksyjnej samotransformacji w świecie coraz doskonalszej predykcji. Człowiek nie powinien zostać zamknięty ani w ograniczeniach biologii, ani we własnej algorytmicznie zrekonstruowanej przeszłości; technologia ma zwiększać przestrzeń możliwego, a nie narzucać, którą konkretną możliwość należy realizować. Drugim postulatem jest zachowanie pluralizmu poznawczego. Systemy powinny wspierać społeczeństwo w dążeniu do konsensusu tam, gdzie dowody są silne, a jednocześnie chronić prawo do niezgody w obszarach, gdzie wiedza pozostaje otwarta lub problem dotyczy wartości. Trzecim jest dystrybucja sprawczości. Wzrost inteligencji organizacji nie może automatycznie oznaczać uszczuplenia realnej autonomii jej członków. Technologie augmentacyjne powinny być oceniane przez pryzmat tego, czy korzyści z władzy poznawczej są dostępne szeroko, czy koncentrują się wyłącznie w centrum instytucji.

Czwartym postulatem jest odpowiedzialność historyczna. System powinien uczyć się na błędach, a instytucja ponosić konsekwencje pozwalające internalizować koszt pomyłek. Sama pamięć techniczna nie wystarczy, jeśli struktura bodźców zachęca do powtarzania tego samego ryzyka. Piątym jest konstytucjonalizacja infrastruktury kognitywnej. Człowiek musi posiadać prawa wobec systemów, od których funkcjonalnie zależy: prawo do wiedzy, kontroli, przeniesienia danych, odwołania i bezpiecznego wyjścia. Wraz ze wzrostem znaczenia AI takie uprawnienia mogą stać się tak fundamentalne dla realnej sprawczości, jak dziś dostęp do komunikacji, edukacji czy usług publicznych. Szóstym jest zachowanie otwartości antropologicznej. Nie wiemy, jak głęboko przyszłe pokolenia zechcą integrować się z technologią i nie powinniśmy narzucać jednej, obowiązkowej wizji „dobrego człowieka przyszłości". Polityka powinna chronić równość moralną osób o różnym stopniu augmentacji.

Siódmym wreszcie jest epistemiczna pokora wobec samej futurologii. Historia technologii pełna jest zarówno niedoszacowanych przemian, jak i efektownych prognoz, które nigdy się nie zmaterializowały. Koncepcje takie jak Metamind, AGE, podmiot wielopodłożowy czy świadoma AI powinny inspirować badania, ale nie mogą być traktowane jako nieuchronne etapy rozwoju. Dopiero po takim uporządkowaniu można odpowiedzieć na pierwotne pytanie artykułu bez popadania w technologiczną mitologię. Poza AGI znajduje się bowiem nie tyle „bardziej

inteligentna AI", co możliwość powstania bardziej złożonego systemu inteligencji, którego elementami będą człowiek, maszyna, pamięć, instytucja, prawo i kultura. Jego dobroć nie będzie wynikała automatycznie z mocy obliczeniowej, lecz zależeć będzie od sposobu organizacji relacji między tymi elementami.

Największym błędem byłoby zatem czekanie na AGI, by dopiero wtedy rozpocząć debatę o świecie „po AGI". Wiele mechanizmów tej przyszłości powstaje już teraz. Cognitive offloading, agentowość, pamięć systemów, personalizacja, syntetyczne relacje, koncentracja infrastruktury i przebudowa organizacyjna zachodzą bez potrzeby rozstrzygania, czy konkretny model osiągnął już poziom AGI. Era post-AGI może zatem rozpocząć się funkcjonalnie przed samym AGI; społeczeństwo może przechodzić do logiki hybrydowej, podczas gdy badacze wciąż spierają się o definicję generalności. Jest to jeden z najważniejszych paradoksów całej problematyki.

Dlatego pytanie „kiedy AGI?" należy uzupełnić pytaniem: „kiedy nasza zależność od AI stanie się na tyle głęboka, że zmieni instytucjonalną definicję normalnego ludzkiego funkcjonowania?". Ten drugi próg może mieć większe znaczenie społeczne niż pierwszy. Nie musi on nastąpić w jednym konkretnym roku ani zostać oficjalnie ogłoszony przez żadne laboratorium.

Od artefaktu do partnera: etyka odwracalności i sprawczości

Scenariusz ten może się ziścić, gdy student bez AI stanie się wyjątkiem, przedsiębiorstwo pozbawione agentów przestanie być konkurencyjne, administracja przyjmie automatyczną mediację za standard, a osobista pamięć cyfrowa stanie się elementem codziennych zobowiązań. Wtedy Hybrydowa Ludzkość będzie już praktyką, zanim jeszcze stanie się ideologią. Właśnie dlatego najważniejszy obowiązek obecnego pokolenia nie polega na próbie przewidzenia każdej przyszłej technologii, lecz na budowaniu reguł zdolnych zachować sprawczość w warunkach technologicznej nieprzewidywalności. Interoperacyjność, prawa podstawowe, pluralizm, przenośność, audyt, odpowiedzialność i instytucjonalna redundancja są wartościowe właśnie dlatego, że działają niezależnie od tego, która z możliwych przyszłości się urzeczywistni. Można to porównać do projektowania opcji realnych w ekonomii: w warunkach wielkiej niepewności kluczowe jest zachowanie możliwości zmiany kursu. Społeczeństwo, które zbyt wcześnie zamknie się w jednym modelu infrastruktury, traci opcjonalność. Z kolei społeczeństwo zachowujące pluralizm technologiczny i prawny może dostosować się do informacji, których dzisiaj nie posiada. Reversibility jest zatem wartością cywilizacyjną. Nie wszystkie działania da się cofnąć, ale architektura systemów powinna unikać niepotrzebnej nieodwracalności.

Dotyczy to danych, infrastruktury, praw, zależności i modeli wychowania. Im większa niepewność co do długoterminowych skutków, tym cenniejsza staje się możliwość korekty. Takie podejście nie jest konserwatyzmem technologicznym; przeciwnie – umożliwia eksperymentowanie bez konieczności trwałego uzależnienia całego społeczeństwa od pierwszej przyjętej architektury. Odwracalność jest warunkiem odpowiedzialnej innowacji.

Wszystkie te argumenty prowadzą do jednego obrazu. Ludzkość nie stoi przed prostym wyborem między zachowaniem „czystej" biologicznej natury a stopieniem się z maszyną. Taka dychotomia nigdy nie opisywała rzetelnie ludzkiej historii. Język, pismo, instytucje i technologia zawsze były zewnętrznymi składnikami ludzkiej zdolności poznawczej. AI zmienia jednak stopień autonomii i adaptacyjności tych elementów. Pismo nie odpowiada autorowi, książka nie przekształca treści pod jego profil psychologiczny, a notes nie wysyła samodzielnie wiadomości.

AI potrafi to wszystko. Właśnie dlatego obecna przemiana jest jakościowo istotna, mimo historycznej ciągłości człowieka jako istoty technologicznej. Można ją określić jako przejście od artefaktu poznawczego do partnera infrastrukturalnego. Partner nie musi być osobą ani posiadać świadomości; wystarczy, że pełni własną dynamiczną rolę przyczynową we wspólnym procesie. To właśnie takie partnerstwo stanie się najprawdopodobniej podstawowym tworzywem Hybrydowej Ludzkości. Jego formy będą zróżnicowane: od prostych asystentów po sieci agentów, od medycznej augmentacji po organizacyjne Metaminds, od relacji edukacyjnych po neurointerfejsy.

Nie ma jednej odpowiedzi na pytanie, która z tych form jest właściwa. Istnieją jednak kryteria pozwalające je oceniać: czy zwiększają zdolność do działania, czy zachowują możliwość krytyki, czy sprawiedliwie rozdzielają korzyści, czy pamiętają własne błędy, czy można je opuścić oraz czy podlegają prawom i procedurom odpowiednim do ich mocy. Właśnie te kryteria tworzą etykę Hybrydowej Ludzkości. Jej celem nie jest obrona człowieka przed każdą integracją, lecz ochrona warunków, dzięki którym integracja pozostaje jego rozszerzeniem, a nie mechanizmem dominacji. Wówczas pojęcie „posthumanizmu" można odczytać inaczej: nie jako koniec człowieka, lecz jako koniec naiwnego obrazu człowieka jako izolowanego, samowystarczalnego umysłu. To przejście do bardziej realistycznej antropologii, w której podmiot istnieje poprzez relacje biologiczne, społeczne i techniczne.

Taka antropologia nie osłabia odpowiedzialności – wręcz przeciwnie. Skoro działanie powstaje w sieci, trzeba projektować tę sieć tak, aby odpowiedzialność nie ginęła między jej elementami. Skoro poznanie jest rozproszone, należy chronić strukturę źródeł. Skoro pamięć może być zewnętrzna, trzeba chronić prawo do pamięci. Skoro sprawczość może być współrealizowana, trzeba pilnować, aby człowiek nie pozostawał z winą za decyzję, której faktycznie nie kontrolował. W ten sposób interdyscyplinarna analiza zatacza pełne koło.

Kognitywistyka pokazała, że umysł nie jest doskonałym racjonalnym procesorem. Neuronauka ujawniła jego ucieleśnienie i plastyczność. Psychologia społeczna wskazała na znaczenie wpływu i relacji. Filozofia umysłu nauczyła odróżniać funkcję od fenomenalności. Epistemologia społeczna zwróciła uwagę na zależność wiedzy od instytucji. Ekonomia odsłoniła wagę bodźców, kosztów i własności. Prawo zaś dostarcza procedur odpowiedzialności.

Przyszłość inteligencji jako historia systemów i relacji

Socjologia dowodzi, że technologia staje się normą poprzez praktykę, natomiast teoria systemów uczy, iż właściwości całości zależą od architektury relacji między jej elementami. Żadna z tych dyscyplin osobno nie rozwiązuje problemu Hybrydowej Ludzkości; razem jednak pokazują, dlaczego redukcja przyszłości do jednej skali AGI jest niewystarczająca. Właśnie dlatego finał tej analizy nie powinien proklamować nadejścia pojedynczego rozwiązania technicznego. Konkluzja, najbardziej odporna z naukowego punktu widzenia, jest bardziej wymagająca: przyszłość inteligencji będzie prawdopodobnie historią systemów, a nie tylko modeli. Kluczowe właściwości mogą wyłaniać się na poziomie sprzężenia człowiek-AI, zespołu, organizacji lub całej infrastruktury społecznej. Nie mamy podstaw, by twierdzić, że systemy te utworzą jedną świadomość. Mamy natomiast wystarczające przesłanki, by oczekiwać, że będą posiadały zbiorową pamięć, zdolność adaptacji, koordynację i przewagę funkcjonalną przekraczającą możliwości pojedynczych elementów.

To wystarczy, aby traktować je jako nową klasę przedmiotów badań. AGE może w tej perspektywie odegrać rolę wartościowej prowokacji teoretycznej; przypomina ona, że czas i konsekwencja są pomijane, gdy inteligencję mierzy się wyłącznie statycznymi benchmarkami. Przyszła wartość naukowa tego podejścia zależy jednak od tego, czy zostanie ono zoperacjonalizowane bez nieuzasadnionego utożsamiania pamięci z qualia. Fused Mind wskazuje, że granice funkcjonalnej kognicji mogą wykraczać poza organizm, lecz jego wartość zależy od wyraźnego rozróżnienia funkcjonalnej integracji i wspólnej świadomości. Supersociety sugeruje, że sztuczna inteligencja może stać się warstwą zbiorowej koordynacji – pod warunkiem, że nie zamieni się w metaforę legitymizującą podporządkowanie jednostki całości. Explanation Engine przypomina z kolei, że przewidywanie nie powinno być ostatecznym celem poznania; jego wartość zależy od pluralizmu, krytyki, testowalności i wyraźnego oddzielenia modelu od dowodu. Hybrydowa Ludzkość łączy wszystkie te pojęcia, pozostając jednak przede wszystkim nazwą problemu, a nie gotowym rozwiązaniem.

Pytanie brzmi: jak stworzyć cywilizację, w której systemy sztuczne radykalnie zwiększają ludzką i zbiorową zdolność poznawczą, nie powodując jednocześnie zaniku zdolności do samodzielnej krytyki, zmiany celów i demokratycznej kontroli nad infrastrukturą? Nie da się na to odpowiedzieć wyłącznie technologią – potrzebujemy instytucji. Nie da się odpowiedzieć samym prawem – potrzebujemy kompetencji ludzi. Nie wystarczy edukacja – potrzebujemy architektury rynku i infrastruktury. Nie wystarczy filozofia – potrzebujemy danych empirycznych o rzeczywistym wpływie długotrwałej współpracy. Te zależności nie są argumentem za bezradnością, lecz dowodem na konieczność interdyscyplinarnego governance, dokładnie takiego, jakiego wymaga system, którego konsekwencje przenikają wiele poziomów rzeczywistości równocześnie.

Przejście od definicji AGI do projektowania relacji hybrydowej

Kluczowym problemem nie jest bowiem to, czy dany model uzyska etykietę AGI. Choć będzie ona istotna dla nauki, rynku i regulacji, historia cywilizacji rzadko respektuje sztywne granice definicyjne. Ludzie zaczną przebudowywać swoje życie wokół systemów wystarczająco użytecznych na długo przed oficjalnym ogłoszeniem przełomu. Przyszłość nadejdzie zatem nie jako deklaracja laboratorium, lecz jako zmiana normalności. Normalne stanie się konsultowanie z AI przed podjęciem decyzji, korzystanie z zewnętrznej pamięci czy delegowanie części zadań agentom. Normalne stanie się współtworzenie tekstów i projektów oraz współpraca z syntetycznym partnerem poznawczym. Wówczas sprawą pierwszoplanową będzie świadomość, że ta nowa normalność jest konstrukcją instytucjonalną, która wciąż może podlegać wyborowi. Główny kierunek analizy prowadzi więc do pytania nie tylko o to, czym stanie się AI, lecz o to, jak człowiek ewoluuje jako jej partner i czy zachowa sprawczość w świecie porowatej granicy między biologicznym i technicznym poznaniem.

Po przeprowadzonej analizie pytanie to można sformułować jeszcze precyzyjniej. Nie wystarczy pytać, co sztuczna inteligencja będzie umiała zrobić; trzeba pytać, których czynności człowiek przestanie dzięki niej wykonywać, jakie nowe kompetencje w zamian rozwinie i kto będzie kontrolował infrastrukturę tej wymiany. Nie wystarczy pytać o to, czy AI będzie posiadała doświadczenie; należy zapytać, jak pamięć i konsekwencje zostaną wpisane w system oraz czy jego uczenie obejmie również szkody ponoszone przez ludzi. Zamiast pytać, czy powstanie superinteligencja, trzeba zapytać, czy funkcjonalnie superinteligentne organizacje będą podlegały prawu, pluralizmowi i demokratycznej kontroli. Nie wystarczy pytać, czy człowiek pozostanie człowiekiem; kluczowe jest to, jakie warunki pozwolą jednostce zmieniać własne technologiczne

rozszerzenia bez utraty godności, praw i ciągłości życia. I wreszcie – nie wystarczy pytać, co znajduje się poza AGI.

Należy zapytać, kim stanie się ludzkość, gdy inteligencja przestanie być właściwością przypisaną do jednej głowy lub maszyny, a stanie się procesem przepływającym przez biologiczne umysły, modele, pamięć, instytucje i sieci. Właśnie to jest właściwy problem Hybrydowej Ludzkości. Nie przyszłość maszyny czy człowieka w izolacji, lecz przyszłość relacji, która stopniowo staje się częścią obu tych historii. Rozstrzygnięcie tej kwestii nie będzie wynikiem jednego odkrycia naukowego ani pojedynczej decyzji prawnej. Będzie efektem tysięcy projektów technicznych, standardów, regulacji, modeli biznesowych i codziennych wyborów użytkowników, które z czasem zestalą się w infrastrukturę. Dlatego teraźniejszość ma znaczenie większe, niż sugeruje pozornie futurystyczny charakter problemu. Projektujemy bowiem nie tylko narzędzia, ale warunki, w których następne pokolenia będą uczyły się rozpoznawać, gdzie kończy się ich własny osąd, a zaczyna sugestia systemu – i czy ta granica musi być w ogóle wyznaczana w ten sam sposób co dzisiaj. Być może najgłębsza przemiana polega nie na tym, że maszyna nauczy się myśleć jak człowiek, lecz na tym, że człowiek nauczy się myśleć w systemie, którego część nie jest ludzka.

Świadome projektowanie sprzężeń jako ochrona sprawczości

Wówczas osią sporu nie będzie już prosta opozycja biologii i technologii, lecz jakość sprzężenia. Pytanie brzmi: czy zwiększa ono wiedzę bez monopolizacji prawdy, sprawczość bez nieodwracalnej zależności, pamięć bez zamrożenia tożsamości, wygodę bez atrofii kompetencji, koordynację bez likwidacji pluralizmu oraz moc bez zaniku odpowiedzialności? W tym sensie można nadać nowe znaczenie pojęciu „instynktu AI". Najważniejszym „instynktem" przyszłego społeczeństwa nie powinna być ani ślepa wiara w rekomendacje algorytmów, ani automatyczny sprzeciw wobec nich. Powinna nim być metapoznawcza zdolność rozpoznawania właściwego rodzaju relacji z systemem: kiedy mu zaufać, kiedy go zweryfikować, kiedy delegować zadania, a kiedy odzyskać bezpośrednią kontrolę i nie pozwolić, by system sam sformułował problem, który ma później rozwiązać.

Jeśli uda się rozwinąć tę zdolność na poziomie jednostek i instytucji, Hybrydowa Ludzkość nie musi oznaczać zaniku człowieka. Może stać się rozszerzeniem tego, co potrafimy odkryć, stworzyć i skoordynować, przy jednoczesnym zachowaniu prawa do krytyki własnych rozszerzeń. Jeżeli jednak zabraknie tej kompetencji, żadna wroga superinteligencja nie będzie potrzebna, by osłabić ludzką sprawczość. Wystarczy infrastruktura na tyle wygodna, że przestaniemy zauważać, iż część naszego środowiska poznawczego ma własnych właścicieli, kryteria optymalizacji i historię decyzji.

Właśnie dlatego źródłowy "drift" jest bardziej realistycznym ostrzeżeniem niż popkulturowy Skynet. Największa transformacja może dokonać się bez jednego dramatycznego momentu – poprzez sumę decyzji, w których delegacja wydaje się lokalnie racjonalna i społecznie niewidoczna. Odpowiedzią na ten dryf nie jest lęk, lecz świadome projektowanie sprzężeń. Nie jest nią obrona człowieka jako biologicznie niezmiennego artefaktu, lecz ochrona człowieka jako podmiotu zdolnego do nauki, zmiany, relacji i refleksyjnego wyznaczania własnych celów. Nie chodzi o zatrzymanie AI na poziomie narzędzia, lecz o stworzenie reguł, dzięki którym narzędzie może stać się partnerem, nie zyskując przy tym niekontrolowanej dominacji nad środowiskiem poznawczym drugiej strony. Nie jest nią również nadzieja, że nowa technologia rozwiąże problemy stworzone przez poprzednie; odpowiedzią jest rozwijanie instytucjonalnej zdolności do korekty kursu. To właśnie w tej zdolności kryje się najbardziej dojrzały odpowiednik doświadczenia – zarówno dla jednostki, jak i systemu społecznego.

Doświadczenie nie polega bowiem na samym przechowywaniu przeszłości, lecz na zdolności do stawania się innym dzięki jej właściwej interpretacji. Jeśli Hybrydowa Ludzkość ma osiągnąć formę zbiorowej dojrzałości, będzie nią nie nieomylność, lecz umiejętność przekształcania błędów w lepszą architekturę przyszłych działań. W takim ujęciu prawdziwa granica poza AGI nie przebiega między inteligencją ludzką a sztuczną.

Przyszłość po AGI jako cywilizacyjna architektura wspólnego poznania

Wybór biegnie pomiędzy systemami, które jedynie zwiększają swoją moc, a tymi zdolnymi do podnoszenia jakości własnej odpowiedzialności. Jeśli rozwój technologiczny pójdzie wyłącznie pierwszą drogą, otrzymamy coraz większą sprawność w optymalizacji przy jednoczesnym braku teorii tego, co właściwie powinno być optymalizowane. Jeżeli jednak równolegle rozwinie się druga ścieżka, sztuczna inteligencja może stać się elementem znacznie dojrzalszej kognicji cywilizacyjnej: takiej, która pamięta, że każda predykcja jest jedynie modelem; wie, że niepewność stanowi immanentną część wiedzy; zachowuje alternatywy i uczy się na własnych błędach; potrafi powstrzymać działanie oraz chroni człowieka nie przed samą technologią, lecz przed utratą możliwości współdecydowania o kierunku własnej ewolucji.

Ostatecznie pytanie „co poza AGI?” nie dotyczy kolejnej litery alfabetu technologicznego. Jest to pytanie o nowy poziom organizacji poznania. Możliwe, że największa inteligencja przyszłości nie zamieszka ani w ludzkiej czaszce, ani w pojedynczym centrum danych, lecz będzie wyłaniać się z relacji pomiędzy mózgami, modelami, pamięcią, kulturą i instytucjami. Nie oznacza to powstania

jednego globalnego umysłu, lecz możliwość zbudowania cywilizacji, której proces poznawczy jest znacznie bardziej zintegrowany, aktywny i świadomy historycznie niż wszystko, co znały wcześniejsze społeczeństwa. Właśnie dlatego fundamentalnym problemem przyszłości nie jest kwestia tego, czy człowiek wygra z maszyną – w systemie hybrydowym samo pojęcie takiego zwycięstwa traci sens. Człowiek może bowiem korzystać z przewagi AI i równocześnie stać się zależny od jej właściciela; zwiększać produktywność kosztem przenośnych kompetencji; poszerzać pamięć, tracąc zdolność zapominania; otrzymywać lepsze odpowiedzi, tracąc jednocześnie kontakt z procesem ich uzasadniania. Ocena przyszłości musi zatem obejmować całość bilansu sprawczości, a nie pojedynczy parametr sukcesu. Taki bilans nigdy nie zostanie definitywnie zamknięty, gdyż nowe technologie będą nieustannie zmieniać kryteria oceny. To właśnie dlatego governance powinno pozostać refleksyjne: zdolne nie tylko do regulowania systemu, ale i do weryfikacji, czy przyjęte wcześniej pojęcia kontroli, prywatności i odpowiedzialności pozostają aktualne.

Hybrydowa Ludzkość będzie zatem musiała uczyć się nie tylko obsługi sztucznej inteligencji, lecz przede wszystkim samej siebie w jej obecności. Być może właśnie to jest najgłębszy sens AGE, jaki można obronić bez przekraczania granic obecnej wiedzy: przyszłość inteligencji nie jest historią mnożenia poprawnych odpowiedzi, lecz dziejami systemów, które poprzez czas, relację i konsekwencje zmieniają sposób definiowania problemów, własnych ograniczeń oraz odpowiedzialności za wynik. Wtedy pytanie o naturę kognicji ludzkiej i sztucznej otrzymuje odpowiedź, która nie redukuje jednej do drugiej. Kognicja człowieka jest biologiczna, ucieleśniona, historyczna, społeczna i kulturowa. Kognicja AI jest konstruowana, obliczeniowa, skalowalna i coraz sprawniejsza w długotrwałym działaniu w świecie. Ich połączenie nie prowadzi automatycznie do jednolitego umysłu, lecz może stworzyć nową klasę systemów poznawczych, których możliwości i zagrożenia rodzą się w samej relacji.

AGI jest jednym z progów tej historii. AGE może być pytaniem, które należy zadać po przekroczeniu ograniczeń myślenia wyłącznie w kategoriach kompetencji. Hybrydowa kognicja jest mechanizmem, dzięki któremu technologiczna możliwość przenika do codziennego życia, natomiast Hybrydowa Ludzkość stanowi społeczną konsekwencję skali tego procesu. A „poza AGI” znajduje się nie tyle maszyna, co cywilizacja ucząca się funkcjonować jako wielopoziomowy system biologiczno-sztucznego poznania.

Hybrydowa Ludzkość jako świadomie projektowana architektura wspólnego poznania

Jej powodzenie nie będzie mierzone jedynie liczbą rozwiązanych problemów. Będzie mierzone tym, czy pośród rosnącej mocy potrafi zachować człowieka zdolnego do postawienia pytania, którego system sam by nie zadał; obywatela gotowego zakwestionować rekomendację infrastruktury; naukowca potrafiącego rozpoznać słabość dominującego modelu; dziecko zdolne uczyć się bez natychmiastowej odpowiedzi; organizację pamiętającą własne błędy oraz wspólnotę, która potrafi odmówić technologicznie możliwego działania, gdy jego cena narusza wartości, których nie wolno redukować do funkcji optymalizacyjnej. Dopiero wtedy inteligencja stanie się czymś więcej niż wzrostem mocy. Doświadczenie przestanie być tylko archiwum przeszłości, a integracja – zwykłą zależnością.

Dopiero wtedy superinteligencja, jeśli nadejdzie w postaci maszyny, organizacji, Metamind lub sieci wielu podmiotów, będzie mogła zostać osadzona w cywilizacji posiadającej nie tylko środki poznawcze, lecz także instytucjonalną zdolność odpowiedzialnego posługiwania się nimi. Na tej granicy pytanie „co stanie się z AI?" ustępuje miejsca kwestii ważniejszej z perspektywy nauki o człowieku: jaką formę sprawczości, odpowiedzialności i wspólnoty zbuduje gatunek, który po raz pierwszy tworzy nie tylko narzędzia pomagające mu myśleć, lecz systemy aktywnie uczestniczące w tym, skąd biorą się jego myśli, decyzje i obrazy możliwej przyszłości? Odpowiedź pozostaje otwarta, gdyż nie jest wyłącznie prognozą technologiczną, lecz trwającym procesem społecznym.

Dlatego najbardziej odpowiedzialnym stanowiskiem nie jest ani wiara w to, że technologia nieuchronnie poprowadzi człowieka ku wyższej formie istnienia, ani przekonanie, że każda głębsza integracja oznacza jego degradację. Nauka pozwala dziś sformułować wniosek zarazem skromniejszy i donioślejszy: architektura relacji człowieka z AI będzie jednym z czynników współokreślających przyszłą architekturę samego człowieczeństwa. Poza AGI nie czeka gotowy następca Homo sapiens, lecz przestrzeń decyzji o tym, jak przepuszczalna stanie się granica między człowiekiem a jego technologią, jakie prawa zostaną zachowane po obu jej stronach i które kompetencje pozostaną w człowieku, a które w infrastrukturze. Kluczowe będzie to, czy wzrost zbiorowej inteligencji zostanie powiązany z równie szybkim rozwojem zbiorowej odpowiedzialności. Od odpowiedzi na te pytania zależeć będzie, czy Hybrydowa Ludzkość okaże się projektem zwiększającym zdolność człowieka do rozumienia świata i samego siebie, czy tylko niezwykle wydajną infrastrukturą podejmowania decyzji, których coraz mniej ludzi potrafi samodzielnie zrekonstruować.

Między tymi możliwościami nie leży jedna granica technologiczna, lecz cały obszar prawa, edukacji, kultury, nauki, polityki, ekonomii i codziennych praktyk. To w nich ludzie dopiero uczą się, co warto pozostawić maszynie, co współtworzyć, a czego – właśnie dlatego, że może zostać łatwo zautomatyzowane – trzeba świadomie uczyć się samodzielnie. Tak domyka się zasadniczy paradoks przyszłości inteligencji: im więcej człowiek będzie mógł przekazać systemom sztucznym, tym ważniejsza stanie się wiedza o tym, czego przekazywać nie powinien. Im doskonalsza będzie zewnętrzna pamięć, tym cenniejsza umiejętność wyboru tego, co zachować, a co pozwolić sobie zapomnieć.

Im trafniejsze będą predykcje, tym ważniejsze prawo do życia w sposób, którego własny model nie przewidział. Im silniejsza stanie się zbiorowa kognicja, tym większego znaczenia nabierze ochrona indywidualnego sprzeciwu. Im bardziej inteligentna będzie infrastruktura, tym pilniejsza potrzeba instytucji zdolnych powiedzieć jej „nie". Nie jest to sprzeciw wobec inteligencji, lecz warunek jej cywilizacyjnej dojrzałości. Dojrzałość nie polega bowiem na zdolności wykonania wszystkiego, co możliwe, lecz na rozpoznaniu, które możliwości warto urzeczywistnić, komu mają służyć i jakie granice należy zachować, aby przyszłe pokolenia nadal mogły dokonywać własnych wyborów. Jeśli w przyszłości pojawi się coś, co zasłuży na nazwę Metamind, Supersociety lub wielopodłożowej inteligencji, jego najwyższym sprawdzianem nie będzie szybkość obliczeń czy planowania. Będzie nim to, czy system o skali poznawczej przekraczającej ludzką potrafi zostać zorganizowany tak, by nie niszczył warunków wolności, pluralizmu i odpowiedzialności, dzięki którym jego inteligencja w ogóle może służyć wspólnocie. Dopiero wtedy odpowiedź na pytanie „co poza AGI?" przestanie być futurologiczną zagadką.

Będzie nią Hybrydowa Ludzkość rozumiana nie jako podporządkowanie biologii maszynie ani maszyny biologii, lecz jako świadomie projektowana, krytycznie kontrolowana i zdolna do samokorekty architektura wspólnego poznania – taka, w której wzrost mocy systemu idzie w parze ze wzrostem jego odpowiedzialności, a poszerzanie granic inteligencji nie usuwa prawa człowieka do współdecydowania o znaczeniu i kierunku własnej przyszłości.

> Paradoksalnie, im doskonalsza stanie się nasza zewnętrzna pamięć i predykcja, tym cenniejsze będzie prawo do życia w sposób, którego żaden model nie przewidział. Prawdziwym sprawdzianem dla przyszłej superinteligencji nie będzie zatem szybkość jej obliczeń, lecz zdolność do współistnienia z człowiekiem bez niszczenia warunków jego wolności. Czy potrafimy zbudować system tak potężny, by jednocześnie posiadał on odwagę i kulturę powiedzenia nam „nie" w imię wartości, których nie da się sprowadzić do funkcji optymalizacyjnej?

Inspiracje

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

Rana Gujral jest amerykańskim przedsiębiorcą, dyrektorem i technologiem w dziedzinie sztucznej inteligencji i obliczeń kognitywnych. Pełnił funkcję dyrektora generalnego Behavioral Signals, firmy zajmującej się sztuczną inteligencją, która rozwija technologie sztucznej inteligencji oparte na emocjach i kognitywnego rozpoznawania mowy, aby analizować sygnały akustyczne pod kątem ludzkich intencji, emocji i wzorców zachowań. Przed pracą w Behavioral Signals, Gujral założył platformę chmurową dla przedsiębiorstw TiZE (później przejętą przez Alchemy) i piastował stanowiska kierownicze w firmach z branży technologii konsumenckich, takich jak Logitech i Cricut. Był aktywnym prelegentem, inwestorem i współautorem publikacji technologicznych na temat etycznego wdrażania i operacyjnego wpływu sztucznej inteligencji. Gujral jest autorem książki „The AI Instinct: The Future of AI and Human Decision-Making" (Wiley), w której sformułował ramy koncepcji „Artificial General Experience" (AGE) w odniesieniu do hybrydowego procesu decyzyjnego i systemów kognitywnych opartych na interakcji człowiek-maszyna.

Rana Gujral *is an American entrepreneur, executive, and technologist in the field of artificial intelligence and cognitive computing. He served as the Chief Executive Officer of Behavioral Signals, an enterprise artificial intelligence company that develops emotion AI and cognitive speech-recognition technologies to analyze acoustic cues for human intent, emotion, and behavioral patterns. Prior to Behavioral Signals, Gujral founded the enterprise cloud platform TiZE (later acquired by Alchemy) and held senior leadership roles at consumer technology companies, including Logitech and Cricut. He has been an active keynote speaker, investor, and contributor to technology publications on the ethical deployment and operational impact of artificial intelligence. Gujral is the author of The AI Instinct: The Future of AI and Human Decision-Making (Wiley), where he formulated the framework of Artificial General Experience (AGE) to address human-machine hybrid decision-making and cognitive systems.*

Behavioral Signals

Literatura uzupełniająca

Artykuły naukowe

Artykuły naukowe autora:

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

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch02

Źródło

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

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch03

Źródło

[3] "The AI Instinct"

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.fmatter

Źródło

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

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch04

Źródło

[5] "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: 632fec2f-5181-47d3-91a4-4bf7791ea347