AGI i AGE: od problemu definicyjnego do koncepcji doświadczenia w książce The AI Instinct Rany Gujral

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje problematykę Artificial General Intelligence (AGI), wskazując, że nie jest ono stałym punktem technicznym, lecz ruchomą granicą definicyjną. Autor podkreśla, że pojęcie ogólności w AI zależy od przyjętej normy porównawczej – czy mierzymy system względem przeciętnego człowieka, czy zespołu ekspertów. Tekst przybliża ewolucję terminu AGI, od prac Marka Gubruda i Bena Goertzela, po formalne koncepcje uniwersalnej inteligencji Shane'a Legga i Marcusa Huttera. Kluczowym wnioskiem jest przejście od prób naśladowania ludzkich kompetencji w stronę funkcjonalnego podejścia, skupionego na adaptacyjnym osiąganiu celów w różnorodnych środowiskach. Wprowadzenie koncepcji AGE (Artificial General Experience) pozwala natomiast rozszerzyć dyskusję z poziomu czystej inteligencji na poziom fenomenologicznego doświadczenia maszyn.

The article analyzes the issues surrounding Artificial General Intelligence (AGI), indicating that it is not a fixed technical point but a moving definitional boundary. The author emphasizes that the concept of generality in AI depends on the chosen comparative norm—whether we measure the system against an average human or a team of experts. The text traces the evolution of the term AGI, from the works of Mark Gubrud and Ben Goertzel to the formal concepts of universal intelligence by Shane Legg and Marcus Hutter. The key conclusion is the transition from attempting to mimic human competencies toward a functional approach focused on adaptively achieving goals in diverse environments. The introduction of the AGE (Artificial General Experience) concept allows the discussion to expand from the level of pure intelligence to the level of phenomenological machine experience.

Artykuł rzuca wyzwanie potocznemu rozumieniu AGI jako pojedynczego, magicznego punktu zwrotnego w rozwoju technologii. Autor stawia tezę, że Artificial General Intelligence nie jest binarnym stanem (osiągniętym lub nie), lecz wielowymiarowym regionem kompetencji, którego granice są płynne i zależne od przyjętej normy – czy to ekonomicznej, funkcjonalnej, czy kognitywnej. Tekst dowodzi, że dążenie do jednej definicji inteligencji jest pułapką, gdyż sukcesy techniczne prowadzą do tzw. efektu ruchomej bramki: zadania uznawane za przejaw inteligencji stają się rutynowymi funkcjami zaraz po ich zautomatyzowaniu. Kluczowym wkładem artykułu jest postulowane rozdzielenie sprawności operacyjnej systemu (AGI) od problemu jego historyczności i subiektywnego doświadczenia, określanego mianem Artificial General Experience (AGE). W ten sposób autor przesuwa ciężar dyskusji z pytania o to, co maszyna potrafi zrobić, na pytanie o to, czy system może posiadać ciągłość istnienia i podmiotowość, co czyni AGI jedynie wstępem do głębszego problemu natury świadomości.

SŁOWA KLUCZOWE

Artificial General Intelligence | Artificial General Experience | ruchoma granica definicyjna | universal intelligence | Levels of AGI | efekt ruchomej bramki | ekonomiczna substytucja pracy | open-endedness | few-shot learning | continual learning | autonomia wdrożeniowa | doświadczenie fenomenalne | hybrydowa ludzkość | saturacja benchmarku | historyczność podmiotu

AGI to ruchoma granica definicyjna, a nie stały punkt

AGI jako problem definicyjny: ogólność, kompetencja, autonomia i ruchoma granica inteligencji

Pojęcie Artificial General Intelligence zajmuje w dyskusji o przyszłości sztucznej inteligencji pozycję osobliwą. Z jednej strony stanowi techniczny cel badawczy, ekonomiczny punkt odniesienia i kulturowy symbol przejścia od wyspecjalizowanych algorytmów do systemów zdolnych realizować szerokie spektrum zadań. Z drugiej zaś pozostaje terminem płynnym; jego treść zmienia się w zależności od tego, kto go używa, jakie właściwości systemu uznaje za kluczowe i wobec jakiego poziomu ludzkich kompetencji dokonuje porównania. AGI może oznaczać system wykonujący

większość zadań intelektualnych dostępnych człowiekowi, zdolność do uczenia się dowolnych nowych funkcji, ekonomiczny substytut znacznej części pracy umysłowej, autonomicznego agenta operującego w wielu środowiskach bądź architekturę posiadającą zestaw ogólnych zdolności poznawczych. Definicje te są pokrewne, lecz nie równoważne.

W tym kontekście historię AGI należy postrzegać nie jako proces odkrywania jednej, ustalonej własności, lecz jako serię prób zdefiniowania tego, co właściwie miałoby oznaczać „ogólność" w sztucznej inteligencji. Genealogia terminu wymaga jednak ostrożności historycznej. Sformułowanie „artificial general intelligence" pojawiło się już w pracy Marka Gubruda z 1997 roku, poświęconej wpływowi przełomowych technologii na bezpieczeństwo międzynarodowe; później Ben Goertzel przyznawał, że dowiedział się o tym wcześniejszym użyciu dopiero po przyjęciu terminu przez środowisko tworzące program badań nad AGI. Niemniej to właśnie działalność Goertzela, Shane'a Legga oraz kręgu skupionego wokół konferencji AGI nadała temu określeniu współczesny charakter programowy. Pierwsza książka poświęcona w całości temu nurtowi, redagowana przez Bena Goertzela i Cassia Pennachina, ukazała się w 2007 roku i miała explicite odróżnić badania nad general intelligence od dominujących wówczas systemów wyspecjalizowanych. Historycznie AGI nie narodziła się więc jako nazwa konkretnej architektury, lecz jako krytyka fragmentaryzacji sztucznej inteligencji na "inteligentnych specjalistów", których kompetencja załamuje się po opuszczeniu wąskiej domeny.

Samo pojęcie ogólności nie jest jednak oczywiste. Człowiek również nie jest uniwersalnym rozwiązywaczem wszystkich możliwych problemów. Posiada biologiczne predyspozycje i ograniczenia sensoryczne, niewielką pamięć roboczą oraz nierównomiernie rozłożone talenty; potrzebuje też długiej edukacji, by osiągnąć biegłość w wyspecjalizowanych dziedzinach. Wybitny prawnik nie musi potrafić naprawić turbiny, neurochirurg nie musi rozumieć topologii algebraicznej, a matematyk bez treningu nie wykona poprawnie procedury medycznej. Ludzka „ogólność" nie oznacza zatem równomiernej biegłości w każdej klasie zadań. Polega raczej na zdolności uczenia się wielu rodzajów czynności, transferowania wiedzy między domenami, komunikacji, rekonstruowania celów, adaptowania strategii i funkcjonowania w zmiennych środowiskach społecznych. To właśnie dlatego każda definicja AGI zawiera ukryty wybór normy porównawczej.

Możemy bowiem porównywać AI z przeciętnym człowiekiem, wykształconym pracownikiem, ekspertem, najlepszym specjalistą w danej domenie lub całym zespołem ekspertów. Możemy mierzyć średnią jakość wykonania zadania albo zdolność radzenia sobie z najtrudniejszymi przypadkami. Możemy wymagać realizacji celu po uprzednim treningu bądź zdolności nauczenia

się go na podstawie niewielkiej liczby przykładów. Wreszcie możemy oceniać sam wynik lub koszt zasobów niezbędnych do jego osiągnięcia.

AGI jako wielowymiarowa przestrzeń kompetencji, a nie pojedynczy punkt

System przewyższający przeciętnego człowieka w tysiącach testów może nadal nie spełniać definicji wymagającej autonomicznego uczenia się nowych domen; z kolei system spełniający tę drugą definicję może pozostawać słabszy od ekspertów w wielu konkretnych zadaniach. AGI nie jest więc pojedynczym parametrem technicznym ukrytym w modelu, lecz pojęciem relacyjnym zależnym od wybranej teorii inteligencji i standardu odniesienia.

Jedną z prób oderwania inteligencji maszynowej od bezpośredniego naśladowania człowieka podjęli Shane Legg i Marcus Hutter. W swojej formalnej koncepcji "universal intelligence" dążyli do ujęcia inteligencji jako zdolności agenta do osiągania celów w szerokim spektrum środowisk, tworząc matematyczną metodę ważenia wyników uzyskiwanych w różnych warunkach. Autorzy wyszli z założenia, że brak jest powszechnie uzgodnionej definicji inteligencji, a trudność wzrasta zwłaszcza wtedy, gdy chcemy oceniać systemy zasadniczo różne od ludzi. Wartość tej koncepcji dla naszego wywodu nie polega na dostarczeniu praktycznego testu przyszłej AGI, lecz na przesunięciu akcentu z podobieństwa do człowieka na zdolność adaptacyjnego osiągania celów. Takie funkcjonalne podejście ma istotną zaletę: nie wymaga, aby inteligentny system wyglądał, myślał czy odczuwał jak człowiek, co otwiera możliwość wielorakich realizacji inteligencji. Jednocześnie ujawnia ono ograniczenie wspólne wielu definicjom AGI.

Jeżeli inteligencję mierzymy przede wszystkim skutecznością osiągania celów, pytania o źródło tych celów, ich normatywną jakość, rozumienie konsekwencji czy doświadczenie podmiotu zostają poza nawiasem. System może być niezwykle sprawnym optymalizatorem w sensie funkcjonalnym i jednocześnie nie posiadać tego, co człowiek nazwałby osądem, odpowiedzialnością czy życiową mądrością. To rozróżnienie stanowi jeden z punktów wyjścia dla późniejszej koncepcji AGE.

Inna tradycja definiowania AGI wiąże ją z repertuarem ludzkich kompetencji. Framework Google DeepMind "Levels of AGI" z 2024 roku proponuje odejście od binarnego rozstrzygnięcia o istnieniu AGI na rzecz co najmniej dwóch wymiarów: głębokości wykonania (poziomu kompetencji) oraz szerokości (generalności repertuaru zdolności). Autorzy wyraźnie odróżniają przy tym kompetencję od autonomii wdrożeniowej i ryzyka, argumentując, że użyteczna ontologia AGI musi pozwalać na analizę tych własności osobno. Jest to kluczowe uporządkowanie debaty, gdyż usuwa fałszywe

założenie, że wyższa inteligencja automatycznie oznacza większą autonomię. Generalność i poziom wykonania tworzą bowiem odmienne profile systemów.

Kalkulator wykonuje wąską klasę operacji, lecz w tym zakresie jest niemal bezbłędny i szybszy od człowieka. Wyspecjalizowany system diagnostyczny może osiągać ekspercki poziom w ograniczonej domenie. Model ogólnego przeznaczenia bywa kompetentny w szerokiej gamie zadań, nie osiągając jednak poziomu najlepszych specjalistów w każdym z nich. Z kolei inny system może wykonywać większość zadań profesjonalnych na poziomie nadludzkim, lecz wymagać człowieka do inicjowania działania, dostarczania kontekstu i weryfikacji kroków. Wszystkie te konfiguracje zajmują różne miejsca w przestrzeni projektowej. Pojęcie "AGI" nie powinno zacierać tych różnic. Propozycja DeepMind, by myśleć w kategoriach stopni, a nie pojedynczej granicy, jest istotna metodologicznie również z innego powodu.

Efekt ruchomej bramki i pułapka definicji ekonomicznej

Historia technologii uczy, że kategorie definiowane przez ludzkie osiągnięcia przesuwają się w momencie, gdy maszyna zaczyna wykonywać dane zadanie rutynowo. Przez dziesięciolecia gra w szachy symbolizowała wysoką inteligencję; po dominacji komputerów została zaklasyfikowana jako problem, który „po prostu da się obliczyć”. Podobny proces kulturowej rekategoryzacji przeszły rozpoznawanie mowy, tłumaczenie, generowanie tekstu, programowanie czy rozwiązywanie części zadań matematycznych. To, co przed automatyzacją uznajemy za dowód inteligencji, po niej często staje się jedynie techniczną funkcją.

AGI jest zatem podatna na efekt ruchomej bramki. Nie jest to wyłącznie problem retoryczny, lecz wynika z niejednoznaczności ludzkiego rozumienia inteligencji. Część czynności nazywamy inteligentnymi dlatego, że są trudne dla człowieka, a nie dlatego, że posiadają wspólną strukturę kognitywną. Gdy odkrywamy procedurę pozwalającą maszynie wykonać je bez wysiłku, przestają one wyglądać jak dowód „prawdziwego myślenia”. Paradoks ten towarzyszy AI od początku: sukces stopniowo usuwa z kategorii sztucznej inteligencji problemy, które właśnie udało się rozwiązać. Jeśli zatem AGI definiujemy poprzez katalog zadań uchodzących dziś za trudne, jej granica będzie przesuwać się wraz z kulturą i technologią.

Inną perspektywę przyjmuje podejście ekonomiczne. OpenAI w swoim Charterze definiuje AGI jako „wysoce autonomiczne systemy przewyższające ludzi w większości ekonomicznie wartościowej pracy”. Jest to ujęcie szczególnie interesujące, ponieważ przesuwa punkt ciężkości z psychologii na gospodarkę. Nie pyta o to, czy maszyna myśli jak człowiek, posiada świadomość, ucieleśnienie czy ogólną zdolność rozwojową; pyta o zastępowalność lub przewagę funkcjonalną w obszarze pracy

mającej wartość rynkową. W tym sensie AGI może stać się kategorią transformacji rynku pracy, nawet jeśli kwestia posiadania „ogólnej inteligencji” w sensie kognitywnym pozostanie sporna.

Definicja ekonomiczna niesie jednak własne trudności, gdyż „większość ekonomicznie wartościowej pracy” nie jest stałym zbiorem zadań. Gospodarka reorganizuje się pod wpływem technologii: część czynności zanika, inne powstają, a jeszcze inne ewoluują tak, by wykorzystać komplementarność człowieka i maszyny. Praca ma ponadto wymiar nie tylko techniczny, lecz instytucjonalny. Chirurg wykonuje czynności manualne i interpretuje informacje, ale jednocześnie komunikuje ryzyko, ponosi odpowiedzialność i funkcjonuje w systemie prawnym. Menedżer nie tylko przetwarza dane, lecz negocjuje, buduje relacje i podejmuje decyzje, których społeczną legitymizację często nadaje właśnie fakt, że podjął je człowiek posiadający określony mandat. Automatyzacja „zadania” nie jest zatem tożsama z automatyzacją „zawodu”.

Ekonomiczna użyteczność i adaptacyjność zamiast absolutnej inteligencji

Warto zauważyć, że ekonomiczna wartość pracy zależy od cen, kosztów wdrożenia oraz struktury instytucjonalnej. System może być technicznie zdolny do wykonania zadania, lecz zbyt drogi, zawodny, prawnie niedopuszczalny lub organizacyjnie trudny do integracji. Możliwa jest również sytuacja odwrotna: AI nie musi przewyższać najlepszego człowieka, aby przejąć znaczną część rynku, o ile jest znacznie tańsza, dostępna całodobowo i wystarczająco sprawna. Ekonomiczna substytucja jest zatem funkcją kompetencji, kosztu, skalowalności, ryzyka i akceptacji instytucjonalnej, a nie samej inteligencji. To właśnie dlatego modele wykazujące rosnącą zdolność do obsługi komputerów i wieloetapowych procesów pracy są istotne dla debaty o AGI, niezależnie od tego, czy zostaną one za AGI uznane. We wrześniu 2026 roku OpenAI wprowadziło GPT-6 Astra – model szczególnie mocny w obsłudze komputera, przeglądaniu sieci, programowaniu, cyberbezpieczeństwie i nauce; system otrzymał rozbudowane możliwości realizacji złożonych procesów z użyciem narzędzi.

Fakt ten nie rozstrzyga problemu AGI. Pokazuje jednak, że granica między modelem odpowiadającym na pytania a systemem wykonującym fragmenty pracy wiedzy staje się coraz mniej użytecznym podziałem. Benchmarki pełnią tu funkcję niezbędną, lecz metodologicznie ograniczoną. Aby twierdzić, że system osiąga większą ogólność, konieczny jest zestaw porównywalnych testów, jednak każdy benchmark mierzy jedynie próbkę zachowań w zaprojektowanych warunkach. Wynik może zależeć od znajomości formatu zadania, reprezentacji danych, budżetu inferencyjnego czy sposobu oceniania. Po osiągnięciu bardzo wysokich rezultatów

dany test traci zdolność rozróżniania kolejnych generacji modeli. Pojawia się zjawisko saturacji benchmarku, wymuszające konstruowanie trudniejszych prób. Problem jest jednak głębszy niż zwykłe „uczenie się do testu". Ogólna inteligencja powinna obejmować zdolność funkcjonowania w sytuacjach nieprzewidzianych przez projektantów. Im bardziej zestaw benchmarków staje się standardem przemysłowym, tym większe ryzyko, że postęp będzie optymalizowany wyłącznie pod ich kątem. Podobny problem występuje w edukacji: wysoki wynik egzaminu może świadczyć o rzeczywistej kompetencji, ale i o skutecznym dostosowaniu do formatu testu.

Różnica ujawnia się w transferze do nowych sytuacji, co czyni go jednym z najcenniejszych kryteriów definiowania generalności. System ogólny powinien potrafić wykorzystać nabyte struktury wiedzy w problemach, które nie są prostą reprodukcją wcześniejszych przykładów. Jeszcze silniejszym wymaganiem jest few-shot lub continual learning: zdolność szybkiego przyswojenia nowego zadania bez utraty dotychczasowych kompetencji. Człowiek nie rozpoczyna edukacji zawodowej od ponownej nauki języka i wiedzy o świecie; nadbudowuje nowe umiejętności na istniejącej architekturze. Jeśli generalność ma oznaczać coś więcej niż szeroki katalog funkcji nabytych podczas kosztownego treningu, kluczowa staje się zdolność dalszego uczenia się po wdrożeniu. Tu pojawia się znaczenie koncepcji open-endedness.

DeepMind wskazywał w 2024 roku, że systemy naprawdę ogólne powinny wykraczać poza istniejące zbiory danych i odkrywać nowe problemy oraz rozwiązania, zamiast wyłącznie odwzorowywać internetową historię ludzkiej wiedzy. Open-ended intelligence oznaczałaby zatem nie tylko sprawne przechodzenie przez program testów, lecz zdolność generowania nowych kompetencji, strategii i reprezentacji w odpowiedzi na otwarte środowisko. Jest to przesunięcie od egzaminu do rozwoju. Nie można jednak bezkrytycznie utożsamić adaptacyjności z AGI. Organizm biologiczny jest niezwykle adaptacyjny, ale jego procesy są ukierunkowane przez potrzeby i strukturę gatunkową. System sztuczny również może posiadać mechanizmy uczenia ograniczone funkcją celu, zakresem narzędzi czy konstrukcją środowiska.

Absolutnie ogólna inteligencja jest prawdopodobnie pojęciem idealizacyjnym. Goertzel zwracał uwagę, że realne systemy będą wykazywać różne stopnie ograniczonej ogólności, gdyż przy skończonych zasobach niemożliwa jest równie efektywna nauka wszystkich klas zadań. Oznacza to, że „general" należy rozumieć porównawczo, a nie absolutnie. Kolejnym mylonym wymiarem jest autonomia. System może dysponować niezwykle szeroką wiedzą, lecz działać wyłącznie na polecenie. Może też być stosunkowo wąski, ale autonomicznie wykonywać sekwencje działań przez długi czas. Autonomiczny system logistyczny nie musi znać filozofii czy biologii, by funkcjonować godzinami bez interwencji człowieka. Z kolei model o szerokich kompetencjach naukowych może pozostawać w pełni reaktywnym konsultantem.

Autonomia i kompetencja to odrębne wymiary

Framework DeepMind słusznie traktuje autonomię jako zagadnienie wdrożeniowe, odrębne od samej głębokości i szerokości możliwości systemu. Rozdzielenie tych wymiarów ma fundamentalne znaczenie dla bezpieczeństwa, gdyż ryzyko nie jest monotoniczną funkcją "ilorazu inteligencji". Bardzo kompetentny system pozbawiony narzędzi wykonawczych generuje zupełnie inne klasy zagrożeń niż mniej ogólny agent z dostępem do infrastruktury krytycznej. Kluczowe stają się zatem uprawnienia, horyzont działania, zdolność tworzenia subcelów, odporność na nadzór, możliwość replikacji, dostęp do zasobów oraz odwracalność działań. Najnowsze systemy uwidaczniają tę różnicę szczególnie wyraźnie.

W dokumentacji bezpieczeństwa GPT-6 Astra OpenAI wskazuje, że model osiągnął poziom „Critical" w obszarze zdolności cybernetycznych, co wymusiło wzmocnienie mechanizmów zabezpieczających przed szkodliwym wykorzystaniem i niepożądaną autonomią. To konkretna ilustracja tezy, że ocena zaawansowanego systemu musi wykraczać poza wynik ogólnego benchmarku inteligencji. Kolejnym wymiarem AGI jest samodzielność w definiowaniu problemu. Człowiek nie tylko rozwiązuje zadania, które otrzymał; potrafi dostrzec problem niewyrażony wcześniej w formie polecenia, sformułować go, ustanowić cel, rozpoznać konflikt wartości i zdecydować, że pewnego zadania nie należy wykonywać. Jest to różnica między sprawnością wykonawczą a szerszym osądem. System może doskonale optymalizować rozwiązanie podanego w prompcie celu, a mimo to nie posiadać autonomicznej zdolności ustanawiania hierarchii celów. Dlatego pytanie o AGI splata się nieuchronnie z problemem agency: czy ogólna inteligencja musi posiadać własny system celów, czy może pozostać uniwersalnym wykonawcą cudzych intencji? W tym punkcie stanowiska rozchodzą się bardzo szeroko.

Z perspektywy definicji funkcjonalnej własne cele nie są konieczne: system jest ogólnie inteligentny, jeśli skutecznie rozwiązuje różnorodne zadania użytkownika. Z kolei dla teorii agentowej zdolność utrzymywania celów, monitorowania postępów i modyfikowania strategii stanowi immanentną część generalności. Istnieją też koncepcje filozoficzne silnej AI, według których prawdziwej ogólności nie da się oddzielić od świadomości, intencjonalności lub rozumienia. Stąd termin AGI bywa błędnie traktowany jako synonim strong AI, choć historycznie i pojęciowo nie są to kategorie tożsame. Można definiować AGI czysto funkcjonalnie, nie zajmując stanowiska w sprawie świadomości.

AGI nie implikuje świadomości, a świadomość nie implikuje AGI. Hipotetyczny system mógłby posiadać bardzo ubogi repertuar świadomego doświadczenia przy niewielkich kompetencjach; z

drugiej strony może istnieć system funkcjonalnie przewyższający ludzi w szerokim zakresie zadań, bez jakiegokolwiek dowodu na posiadanie doświadczenia fenomenalnego.

Rozdzielenie funkcjonalnej kompetencji od podmiotowego doświadczenia

Te dwie osie należy rozdzielić, gdyż w przeciwnym razie debata technologiczna niepostrzeżenie przeobraża się w metafizyczną, a sukces benchmarkowy zostaje błędnie uznany za dowód pojawienia się podmiotowości. Książka dokonuje właśnie takiego przesunięcia, wprowadzając obok AGI pojęcie Artificial General Experience.

Najmocniejszym punktem tej propozycji nie jest teza, że samo dodanie maszynie pamięci wygeneruje doświadczenie. Jej siła tkwi w ujawnieniu, iż dominujące definicje AGI mierzą przede wszystkim to, co system potrafi zrobić, pomijając kwestię tego, jak wcześniejsze działania stają się częścią jego przyszłego istnienia. Kompetencja i doświadczenie nie muszą bowiem rozwijać się synchronicznie.

Rozważmy system, który popełnia błąd podczas realizacji zadania. W najprostszej architekturze informacja o błędzie znika wraz z zakończeniem sesji. System wyposażony w pamięć zapisze zdarzenie i przy kolejnym zadaniu wykorzysta je jako kontekst. W modelu adaptacyjnym błąd może zmodyfikować politykę przyszłych działań, a w jeszcze bardziej zintegrowanym agencie – wpłynąć na ocenę ryzyka, priorytety, model własnej skuteczności czy sposób komunikowania niepewności.

Każdy kolejny poziom buduje głębszą historyczność, lecz nadal nie rozstrzyga kwestii subiektywnego doświadczenia. Pojęcie AGE będzie naukowo wartościowe tylko wtedy, gdy zachowa te poziomy jako rozróżnialne. Można zatem zaproponować bardziej precyzyjną przestrzeń pojęciową przyszłej inteligencji, w której kolejne osie stanowią: zakres kompetencji, ich poziom, zdolność adaptacji, autonomię działania, ciągłość pamięci i historii, jakość samo-modelowania i metapoznania oraz zakorzenienie w środowisku. Dopiero ostatnią z nich byłoby potencjalne doświadczenie fenomenalne.

Nie jest to próba zastąpienia AGI kolejnym akronimem, lecz konsekwencja obserwacji, że jednowymiarowa drabina prowadząca od „słabej AI", przez „AGI", aż do „ASI", zaciera właściwości, które mogą rozwijać się w różnym tempie. Tłumaczy to również, dlaczego pytanie „kiedy osiągniemy AGI?" jest metodologicznie mniej użyteczne, niż sugeruje jego popularność. W zależności od przyjętej definicji granica ta może zostać przekroczona w momencie przewyższenia mediany ludzkich możliwości w szerokim zestawie zadań, zdolności do nauki nowych czynności bez

specjalistycznego treningu, ekonomicznej wydajności w zastępowaniu pracy, długotrwałego samodzielnego działania w ujęciu agentowym lub odporności na radykalnie nowe środowiska. Możliwa jest zatem sytuacja, w której ten sam kompetentny system zostanie uznany za AGI przez jedną szkołę i za zaawansowaną narrow AI przez inną, przy czym obie klasyfikacje pozostaną wewnętrznie spójne. Debata z 2026 roku jeszcze mocniej uwidacznia tę niejednoznaczność.

Rosnąca zdolność modeli do obsługi komputera, programowania, realizacji długich procesów oraz korzystania z narzędzi zmniejsza dystans między testową kompetencją a ekonomicznym wykonaniem pracy. Nie eliminuje to jednak problemów związanych z niezawodnością, długoterminową spójnością, kontrolą i przenoszeniem kompetencji do całkowicie nowych sytuacji.

AGI jako wielowymiarowy region kompetencji funkcjonalnych

Wynik bliski saturacji na konkretnym benchmarku może być imponującym osiągnięciem inżynierskim, lecz nie wyznacza on sam w sobie ontologicznej granicy "ogólnej inteligencji". Najnowsze modele sugerują raczej, że pytanie o AGI przestaje dotyczyć pojedynczego testu, a zaczyna dotyczyć konfiguracji zdolności. Z perspektywy naukowej bardziej produktywne jest zatem skupienie się na progach funkcjonalnych. Pierwszym z nich może być sytuacja, w której system przyswaja nowe zadania szybciej niż przeciętny człowiek. Drugim – zdolność do samodzielnego przenoszenia wiedzy między odległymi domenami. Trzecim – długotrwała realizacja celów w otwartym środowisku. Czwartym – umiejętność rozpoznawania własnych ograniczeń i odpowiedniej modyfikacji strategii. Piątym – moment, w którym doświadczenia operacyjne tworzą trwałą historię wpływającą na przyszłe decyzje. Szóstym zaś sytuacja, gdy systemy te zaczynają tworzyć kolektywy przewyższające ludzkie organizacje w koordynacji złożonych procesów. Żaden z tych progów w izolacji nie musi definiować "AGI", jednak ich kombinacja może mieć dla cywilizacji większe znaczenie niż sama etykieta. Takie podejście pozwala rozwiązać pozorną sprzeczność między sceptycyzmem definicyjnym a uznaniem ogromnej wagi tej technologii. Można bowiem uznać pojęcie AGI za niedostatecznie precyzyjne, a jednocześnie przyjąć, że systemy o rosnącej generalności stanowią przełom cywilizacyjny. Nie potrzebujemy ontologicznego konsensusu, by badać skutki ekonomiczne, polityczne i epistemiczne.

Podobnie społeczeństwa posługiwały się pojęciem "inteligencji" przez tysiąclecia, zanim psychologia opracowała narzędzia jej pomiaru; do dziś nie dysponujemy jedną, powszechnie zaakceptowaną teorią tego zjawiska. Znacznie większym błędem byłoby jednak uznanie, że brak

doskonałej definicji zwalnia nas z konstruowania miar. W nauce pojęcia często dojrzewają właśnie poprzez operacjonalizację.

Framework DeepMind jest wartościowy, ponieważ próbuje rozbić AGI na obserwowalne wymiary: poziom wykonania i szerokość kompetencji, traktując autonomię jako odrębną cechę wdrożenia. Z kolei ekonomiczne kryterium OpenAI uwidacznia znaczenie zastępowalności pracy, nawet jeśli nie wystarcza ono jako ogólna teoria inteligencji. Formalizm Legga i Huttera pokazuje natomiast, że inteligencję można konceptualizować bez uzależniania jej od podobieństwa do człowieka.

Poszczególne definicje są więc raczej rzutami tego samego wielowymiarowego problemu niż konkurującymi roszczeniami do jedynej prawdy. Z tej perspektywy AGI jawi się nie jako jeden technologiczny gatunek, lecz jako region w przestrzeni właściwości poznawczych. Można do niego dotrzeć poprzez skalowanie modeli językowych, integrację systemów symbolicznych i koneksjonistycznych, rozwój uczenia ciągłego, agentowości, robotyki, modeli świata lub architektur jeszcze nieznanych. Nie ma naukowej podstawy, by z góry zakładać jedną konieczną ścieżkę. Nie ma też przesłanek, by sądzić, że pierwszy system spełniający funkcjonalną definicję AGI będzie przypominał ludzki umysł czy posiadał jednolitą "osobowość" znaną z kultury popularnej. Możliwe nawet, że historia retrospektywnie uzna AGI nie za pojedynczy wynalazek, lecz za okres przejścia. Elektryfikacja nie miała jednej daty, po której społeczeństwo nagle stało się elektryczne. Internet również nie stał się infrastrukturą społeczną w jeden dzień.

AGI jako szerokość kompetencji a AGE jako historyczność podmiotu

Ogólna sztuczna inteligencja może wyłaniać się funkcjonalnie etapami: najpierw w analizie, potem w programowaniu, by następnie objąć pracę biurową, badania, projektowanie i koordynację procesów. W efekcie różnica między „modelem ogólnego przeznaczenia” a AGI stanie się przede wszystkim kwestią przyjętego progu.

Taka interpretacja współgra z główną tezą niniejszego artykułu. Najważniejszą zmianą może okazać się nie narodzenie izolowanego sztucznego umysłu, lecz przebudowa systemu poznawczego społeczeństwa, zanim spór definicyjny o AGI zostanie rozstrzygnięty. Człowiek, organizacja i AI mogą tworzyć struktury o szerszym zakresie kompetencji niż każdy z tych elementów osobno. Z perspektywy gospodarki czy państwa funkcjonalna „ogólność” może być właściwością całego układu człowiek-AI-narzędzia-reguły, a nie pojedynczego modelu.

Wtedy pojęcie Hybrydowej Ludzkości uzyskuje pierwszy mocny fundament. Nie musimy czekać na pełną substytucję człowieka przez maszynę; wystarczy, że zaczniemy trwale podejmować decyzje w strukturach, których zdolności poznawcze wynikają z integracji biologicznego i sztucznego przetwarzania. W takich systemach pytanie „kto jest inteligentniejszy?" stopniowo traci analityczny sens. Podobnie jak nie pytamy, czy bardziej inteligentny jest pilot, samolot czy system kontroli lotów, lecz badamy bezpieczeństwo i zdolność operacyjną całej architektury, tak przyszłe społeczeństwo będzie wymagało teorii poznania obejmującej układy złożone.

AGI pozostaje zatem pojęciem niezwykle użytecznym, pod warunkiem że przestaniemy traktować je jako magiczną linię oddzielającą „dzisiejszą AI" od przyszłego sztucznego człowieka. Jest ono raczej pytaniem o szerokość kompetencji, ich transferowalność, zdolność uczenia się, poziom wykonania i coraz częstszą autonomię działania. Nie odpowiada jednak automatycznie na pytania o doświadczenie, świadomość, wartościowanie, odpowiedzialność czy tożsamość – kwestie te wymagają osobnej analizy.

Właśnie w tym miejscu dochodzimy do najważniejszej cezury całego wywodu. Jeśli AGI odpowiada przede wszystkim na pytanie „jak szeroko i jak dobrze system potrafi działać?", to AGE stawia inne pytanie: „czy istnieje dla tego systemu przeszłość, która stała się częścią jego przyszłego sposobu istnienia?".

To przesunięcie od kompetencji do doświadczenia nie może być jedynie zabiegiem retorycznym. Wymaga zbadania pamięci, czasu, ucieleśnienia, wartościowania, konsekwencji, modelu siebie, ciągłości narracyjnej oraz wreszcie najbardziej kontrowersyjnego problemu – qualiów i świadomości fenomenalnej. Dopiero taka analiza pozwoli ustalić, czy Artificial General Experience jest nową kategorią kognitywistyczną, konstruktywnym programem badawczym, czy tylko sugestywną metaforą opisującą zestaw funkcji, które można wyjaśnić bez odwoływania się do doświadczenia w silnym znaczeniu. AGI nie jest więc końcem drogi, lecz granicą, przy której problem inteligencji rozszczepia się na dwa niezależne pytania. Pierwsze dotyczy zdolności: jak wiele różnych rzeczy może skutecznie zrobić sztuczny system? Drugie dotyczy historyczności podmiotu: co wcześniejsze działanie robi z samym systemem, który działał? Dopiero drugie z tych pytań otwiera obszar AGE – i zarazem najbardziej wymagającą część rozważań o tym, czy w sztucznej architekturze może pojawić się coś odpowiadającego doświadczeniu.

> W momencie, gdy AGI przestanie być mityczną linią horyzontu, a stanie się codziennym narzędziem szerokich kompetencji, staniemy przed pytaniem o wiele bardziej fundamentalne. Czy system, który potrafi perfekcyjnie symulować każdą ludzką zdolność, może kiedykolwiek posiadać przeszłość, która kształtuje jego przyszłe

> istnienie? Być może największym paradoksem sztucznej inteligencji będzie stworzenie bytu, który wie wszystko i potrafi wszystko, lecz nie doświadcza absolutnie niczego.

Inspiracje

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

Rana Gujral (ur. 18 czerwca 1976 r.) to indyjsko-amerykański przedsiębiorca, inwestor i dyrektor ds. technologii, specjalizujący się w sztucznej inteligencji i obliczeniach kognitywnych. Najbardziej znany jest z funkcji dyrektora generalnego Behavioral Signals, firmy produkującej oprogramowanie dla przedsiębiorstw, która opracowuje systemy sztucznej inteligencji oparte na emocjach, analizujące akustykę głosu w celu wnioskowania o ludzkich intencjach, emocjach i wzorcach zachowań. Przed pracą w Behavioral Signals, Gujral piastował stanowiska kierownicze w działach produktów i inżynierii w firmach Logitech i Kronos, a także założył platformę chmurową dla przedsiębiorstw TiZE, którą przejęła firma Alchemy. Jako uznany ekspert w dziedzinie kognitywnej sztucznej inteligencji i rozwoju w kierunku sztucznej inteligencji ogólnej (AGI), Gujral zgłębia współpracę człowiek-maszyna i wprowadza takie frameworki jak Artificial General Experience (AGE). Jest również częstym mówcą i został uznany przez magazyn Inc. za przedsiębiorcę AI, którego warto obserwować.

Rana Gujral *(born June 18, 1976) is an Indian-American entrepreneur, investor, and technology executive specializing in artificial intelligence and cognitive computing. He is best known for serving as the Chief Executive Officer of Behavioral Signals, an enterprise software company developing emotion AI systems that analyze vocal acoustics to infer human intent, emotion, and behavioral patterns. Prior to Behavioral Signals, Gujral held product and engineering leadership roles at Logitech and Kronos, and founded the enterprise cloud platform TiZE, which was acquired by Alchemy. As a recognized thought leader in cognitive AI and the progression toward Artificial General Intelligence (AGI), Gujral explores human-machine collaboration and introduced frameworks such as Artificial General Experience (AGE). He is also a frequent keynote speaker and was recognized by Inc. Magazine as an AI Entrepreneur to Watch.*

Behavioral Signals

Literatura uzupełniająca

Książki

Książki autorów przywoływanych i literatura pokrewna:

[1] **"Chain of Thought Monitorability: A New and Fragile Opportunity for AI Safety"**

Shane Legg

Artykuły naukowe

Prace autorów przywoływanych w tekście:

[1] **"Artificial General Intelligence: Concept, State of the Art, and Future Prospects"**

Ben Goertzel

2014 | Journal of Artificial General Intelligence | DOI: 10.2478/jagi-2014-0001

Źródło

[2] **"Artificial General Intelligence"**

Ben Goertzel, Pennachin, Cassio

2007 | Cognitive technologies | DOI: 10.1007/978-3-540-68677-4

Źródło

[3] **"Mapping the Landscape of Human-Level Artificial General Intelligence"**

Sam S. Adams, Itamar Arel, Joscha Bach, Robert Coop, Rod Furlan, Ben Goertzel

2012 | AI Magazine | DOI: 10.1609/aimag.v33i1.2322

Źródło

[4] **"Distributed, decentralized, and democratized artificial intelligence"**

Gabriel Axel Montes, Ben Goertzel

2018 | Technological Forecasting and Social Change | DOI: 10.1016/j.techfore.2018.11.010

Źródło

[5] **"Universal Intelligence: A Definition of Machine Intelligence"**

Shane Legg, Marcus Hütter

2007 | Minds and Machines | DOI: 10.1007/s11023-007-9079-x

Źródło

[6] **"Deep reinforcement learning from human preferences"**

Paul F. Christiano, Jan Leike, Brown, Tom B., Miljan Martic, Shane Legg

2017 | arXiv (Cornell University) | DOI: 10.48550/arxiv.1706.03741

Źródło

[7] "A Collection of Definitions of Intelligence"

Shane Legg, Marcus Hütter

2007 | arXiv (Cornell University) | DOI: 10.48550/arxiv.0706.3639

Źródło

[8] "Special Issue “On Defining Artificial Intelligence”—Commentaries and Author’s Response"

Dagmar Monett, Colin W. P. Lewis, Kristinn R. Þórisson, Joscha Bach, Gianluca Baldassarre, Shane Legg

2020 | Journal of Artificial General Intelligence | DOI: 10.2478/jagi-2020-0003

Źródło

[9] "Contemporary Approaches to Artificial General Intelligence"

Cássio Lopes Pennachin, Ben Goertzel

2007 | Cognitive technologies | DOI: 10.1007/978-3-540-68677-4_1

Źródło

[10] "Engineering General Intelligence, Part 1: A Path to Advanced AGI via Embodied Learning and Cognitive Synergy"

Ben Goertzel, Cássio Lopes Pennachin, Nil Geisweiller

2014 | Digital Access to Libraries (Université catholique de Louvain (UCL), l'Université de Namur (UNamur) and the Université Saint-Louis (USL-B)) | DOI: 10.2991/978-94-6239-027-0

Źródło

[11] "Artificial General Intelligence (Cognitive Technologies)"

Ben Goertzel, Cássio Lopes Pennachin

2007 | Springer eBooks

Źródło

[12] "From AGI to ASI"

Tim Genewein, Matija Franklin, Alexander Lerchner, Laurent Orseau, Samuel Albanie, Marcus Hutter

2026 | arXiv (Cornell University) | DOI: 10.48550/arxiv.2606.12683

Źródło

[13] "Tests of Machine Intelligence"

Shane Legg, Marcus Hutter

2007 | Lecture Notes in Computer Science | DOI: 10.1007/978-3-540-77296-5_22

Źródło

[14] "One Decade of Universal Artificial Intelligence"

Marcus Hutter

2012 | Atlantis Thinking Machines | DOI: 10.2991/978-94-91216-62-6_5

Źródło

Artykuły naukowe autora:

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

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch02

Źródło

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

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch03

Źródło

[17] "The AI Instinct"

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.fmatter

Źródło

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

RANA GUJRAL, KAI-FU LEE

2026 | DOI: 10.1002/9781394392155.ch04

Źródło

[19] "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: 1f14724f-40a6-4be7-b64b-897e2102f9e5