Hybrydowa kognicja i emergencja w relacji człowiek-AI w świetle koncepcji Boba Burgina



AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Tekst analizuje ewolucję ludzkiego poznania w kontekście integracji z systemami sztucznej inteligencji. Autor, odwołując się do teorii rozszerzonego umysłu Clarka i Chalmersa oraz koncepcji Boba Burgina, argumentuje, że AI przestała być neutralnym narzędziem, a stała się aktywnym elementem systemu poznawczego. W przeciwieństwie do pasywnych nośników pamięci (jak książki czy notatniki), generatywna AI pełni rolę operatora kognitywnego, który aktywnie przekształca informacje i modyfikuje procesy myślowe użytkownika w czasie rzeczywistym. Zjawisko to określa się mianem interaktywnej zewnętrznej kognicji (interactive external cognition), gdzie następuje synergia między biologiczną architekturą mózgu a cyfrową infrastrukturą, co prowadzi do eksternalizacji funkcji poznawczych i zmiany sposobu, w jaki człowiek tworzy i przetwarza wiedzę.

The text analyzes the evolution of human cognition in the context of integration with artificial intelligence systems. Drawing on Clark and Chalmers' theory of the extended mind and Bob Burgin's concepts, the author argues that AI has ceased to be a neutral tool and has become an active element of the cognitive system. Unlike passive memory carriers (such as books or notebooks), generative AI acts as a cognitive operator that actively transforms information and modifies the user's thought processes in real time. This phenomenon is termed interactive external cognition, where synergy occurs between the biological architecture of the brain and digital infrastructure, leading to the externalization of cognitive functions and a change in how humans create and process knowledge.

Artykuł analizuje ewolucję sztucznej inteligencji z roli neutralnego narzędzia do statusu aktywnego komponentu hybrydowego systemu poznawczego. Autor stawia tezę, że AI nie tylko rozszerza ludzkie możliwości pamięciowe i twórcze (external cognition), ale realnie modyfikuje procesy myślowe, tworząc pętlę sprzężenia zwrotnego między człowiekiem a maszyną. Tekst ostrzega przed pułapką poznawczej homogenizacji – sytuacją, w której indywidualny wzrost produktywności odbywa się kosztem zbiorowej różnorodności idei i zaniku czasu niezbędnego na biologiczną inkubację myśli. Kluczowym wyzwaniem staje się zatem wypracowanie nowej sprawczości epistemicznej: umiejętności świadomego zarządzania delegowaniem funkcji poznawczych tak, aby augmentacja nie stała się substytutem autonomii intelektualnej. W tym ujęciu AI jest nowym elementem niszy poznawczej człowieka, który wymusza redefinicję pojęć autorstwa, samodzielności i tożsamości w świecie, gdzie granica między procesami wewnętrznymi a zewnętrznymi staje się płynna.

SŁOWA KLUCZOWE

hybrydowa kognicja | emergencja w relacji człowiek-AI | rozszerzony umysł | eksternalizacja funkcji poznawczych | interaktywna zewnętrzna kognicja | pamięć transakcyjna | sprawczość epistemiczna | poznawcza homogenizacja | myślenie dywergencyjne | cyfrowy hipokamp | uczenie relacji poznawczej | architektura predykcyjna | synergia kognitywna | fiksacja kreatywna

AI jako aktywny element systemu poznawczego

Umysł poza czaszką: hybrydowa kognicja, pamięć zewnętrzna, autorstwo i twórczość człowieka żyjącego z AI

Dotychczasowe rozważania prowadziły przede wszystkim od biologii ku technice: od synaps i pamięci, przez predykcję, ucieleśnienie i świadomość, aż po modele sztucznej inteligencji wykazujące coraz więcej funkcjonalnych analogii do procesów poznawczych. W pewnym momencie kierunek analizy musi jednak zostać odwrócony.

Jeżeli człowiek przez znaczną część dnia korzysta z systemu, który przypomina mu wcześniejsze ustalenia, porządkuje informacje, generuje alternatywy, wyszukuje argumenty, pomaga planować działanie, krytykuje tekst, symuluje rozmówcę i podsuwa język do wyrażenia jeszcze nieukształtowanej myśli, sztuczna inteligencja przestaje być neutralnym przedmiotem znajdującym się poza poznaniem. Zaczyna modyfikować sam sposób, w jaki człowiek pamięta, poszukuje, rozważa i tworzy. Pytanie o świadomość AI staje się wówczas tylko częścią większego problemu. Niezależnie od tego, czy system sztuczny cokolwiek przeżywa, może on już dziś stanowić element procesu poznawczego człowieka. Burgin formułuje tę intuicję radykalnie.

AI zostaje w nim określona jako „eksternalizacja" ludzkiej architektury predykcyjnej – sposób przekroczenia ograniczeń biologicznego mózgu poprzez podłączenie poznania do zewnętrznej infrastruktury energetycznej i pamięciowej. W innym miejscu kultura zostaje nazwana rodzajem "zewnętrznego dysku", którego ograniczeniem jest powolna transmisja międzyludzka, podczas gdy systemy sztuczne miałyby umożliwiać niemal natychmiastową synchronizację doświadczeń. Kierunek tej myśli jest interesujący, lecz termin „eksternalizacja naszej architektury" wymaga ostrożności. Współczesny model językowy nie jest ludzkim mózgiem przeniesionym poza czaszkę. Nie odtwarza mechanicznie jego neurobiologii, homeostazy, rozwoju ani historii ucieleśnienia. Jest raczej nowym artefaktem poznawczym zbudowanym na ogromnym zbiorze wcześniej zeksternalizowanych wytworów ludzkiego poznania – języka, obrazów, kodu i dokumentów – który może następnie zwrotnie oddziaływać na ludzi tworzących kolejne elementy kultury. Pojęcie eksternalizacji ma zatem większą wartość, jeśli zastosować je nie do architektury neuronalnej, lecz do funkcji poznawczych.

Człowiek od tysięcy lat przenosi poza mózg część pamięci, obliczeń i organizacji działania. Pismo pozwala przechowywać komunikat dłużej niż żyje człowiek go pamiętający. Notacja matematyczna umożliwia rozumowania przekraczające pojemność pamięci roboczej. Książka pozwala spotkać się z myślą autora nieobecnego od stuleci. Biblioteka reorganizuje pamięć zbiorową, mapa przenosi reprezentacje przestrzenne na powierzchnię, zegar eksternalizuje pomiar czasu, a komputer pozwala delegować ogromne klasy operacji symbolicznych. AI należy umieścić w tej historii nie jako pierwszy przypadek "umysłu poza mózgiem", lecz jako potencjalnie nowy etap – zewnętrzny artefakt, który nie tylko przechowuje reprezentacje, ale aktywnie je przekształca.

Ta różnica prowadzi bezpośrednio do klasycznej tezy Andy'ego Clarka i Davida Chalmersa o umyśle rozszerzonym. Ich argument z 1998 roku wychodził od funkcjonalnego pytania: jeżeli proces zewnętrzny jest zintegrowany z działaniem osoby w taki sposób, że pełni funkcję podobną do procesu, który bez wahania nazwalibyśmy poznawczym, gdyby znajdował się w mózgu, dlaczego automatycznie wykluczać go z systemu poznawczego tylko dlatego, że przekracza granicę skóry?

Słynny notatnik Otto pełnił dla jego właściciela funkcję trwałego zewnętrznego zasobu pamięciowego.

Współczesny Clark rozwija tę intuicję wprost w odniesieniu do generatywnej AI. Argumentuje, że praktycznym problemem nie jest metafizyczna obrona granicy czaszki, lecz kompetencja w ocenianiu, kiedy zewnętrznemu zasobowi można zaufać i kiedy poznawcze odciążenie jest korzystniejsze od użycia pamięci biologicznej. Istnieją jednak różne stopnie rozszerzenia. Zapisane w notesie hasło jest bierne.

AI jako aktywny operator poznawczy w pętli interakcji

Wyszukiwarka wykazuje większą aktywność, gdyż porządkuje wyniki według algorytmicznej funkcji relewancji. System rekomendacyjny idzie krok dalej, modelując profil użytkownika. Generatywna AI potrafi natomiast przejąć niepełną myśl, rozwinąć kilka możliwych interpretacji, wskazać kontrargumenty, połączyć ją z odległym obszarem wiedzy i zaproponować dalszą sekwencję działania.

Nie jest to już tylko „miejsce, w którym coś pamiętam". AI staje się operatorem na reprezentacjach, a jej odpowiedzi kształtują kolejny stan poznawczy użytkownika. Pozwala to odróżnić external memory od external cognition. Biblioteka przechowuje informacje, których człowiek sam nie pamięta; AI natomiast potrafi przeprowadzać transformacje tych informacji. Jeszcze dalej znajduje się interactive external cognition, w której każda transformacja modyfikuje kolejne pytanie użytkownika, a każde pytanie wpływa na działanie systemu. Powstaje pętla sprzężenia, której rezultat nie musi istnieć w żadnym z elementów przed rozpoczęciem interakcji. Człowiek nie posiadał gotowej myśli, AI nie miała odpowiedzi na konkretny, jeszcze niesformułowany problem – finalna argumentacja wyłania się z procesu iteracyjnego. Właśnie tutaj pojęcie emergencji znajduje zastosowanie na nowym poziomie: mamy system złożony z dwóch komponentów, z których każdy posiada własne ograniczenia.

Człowiek dysponuje ograniczoną pamięcią roboczą, czasem i cierpliwością, ale posiada ucieleśnioną biografię, cele, odpowiedzialność, kontekst społeczny oraz zdolność nadawania znaczenia doświadczeniu. AI oferuje ogromną sprawność w przeszukiwaniu przestrzeni reprezentacji i generowaniu wariantów, lecz jest pozbawiona biologicznych ograniczeń, osobistej historii czy gwarantowanej fenomenologii. Pytanie brzmi, czy ich połączenie może osiągać kompetencje, których żaden z komponentów nie zrealizowałby samodzielnie. Nie należy zakładać odpowiedzi twierdzącej z góry – 1 + 1 nie zawsze daje więcej niż 2. Człowiek może zostać zakotwiczony przez pierwszą sugestię modelu, zbyt łatwo zaufać fałszywej odpowiedzi, utracić motywację do weryfikacji

informacji lub przestać dostrzegać alternatywy niewygenerowane przez system. AI z kolei może dziedziczyć uprzedzenia z danych, nie rozpoznawać lokalnego kontekstu i formułować błędne odpowiedzi z nadmierną pewnością.

Efektem takiej współpracy może być synergia, ale równie dobrze kumulacja lub korelacja błędów. Badania nad współpracą człowieka z AI w medycynie, analizujące dziesiątki wcześniejszych eksperymentów, wskazują, że osiągnięcie trwałego efektu „1 + 1 > 2" pozostaje problemem empirycznym, a nie pewnikiem wynikającym z samego faktu obecności dwóch różnych systemów poznawczych. Do zrozumienia tej relacji przydatna jest teoria poznania rozproszonego Edwina Hutchinsa. W tym ujęciu jednostką analizy nie musi być pojedynczy mózg; system poznawczy może obejmować grupę ludzi, artefakty reprezentacyjne, reguły działania i środowisko. Nawigacja statku nie odbywa się „w głowie" jednego człowieka – jest rezultatem skoordynowanego działania osób, instrumentów, map, procedur i przepływu informacji.

AI jako zewnętrzny partner pamięciowy i jego wpływ na tożsamość

Podobnie współczesne przedsiębiorstwo pamięta więcej niż którykolwiek z jego pracowników. Wiedza jest rozproszona: część drzemie w ludziach, część w dokumentacji i procedurach, a jeszcze inna w oprogramowaniu czy relacjach między rolami. AI wprowadza do takiego systemu element szczególny, gdyż potrafi dynamicznie reorganizować informacje pochodzące od wszystkich pozostałych komponentów. Z tej perspektywy nie trzeba uznawać modelu za autonomiczny umysł, by stwierdzić, że organizacja wspierana przez AI zyskuje inną architekturę poznawczą niż ta pozbawiona tego narzędzia. Zmienia się prędkość wyszukiwania informacji, sposób delegowania analizy, koszt przygotowania pierwszej wersji dokumentu, możliwość symulowania wariantów oraz zakres wiedzy dostępny pojedynczemu pracownikowi. Podobny mechanizm opisuje teoria pamięci transakcyjnej Daniela Wegnera.

W grupie człowiek nie musi wiedzieć wszystkiego, o ile wie, kto posiada daną informację. Para, rodzina czy zespół mogą rozdzielać obszary pamięci: jedna osoba pamięta terminy, druga szczegóły techniczne, trzecia historię kontaktów. System zbiorowy dysponuje większą wiedzą niż którakolwiek z jednostek, ponieważ jego kluczowym składnikiem jest metawiedza o rozmieszczeniu zasobów. Zewnętrzna AI radykalizuje ten układ.

Użytkownik może przestać pamiętać nie tylko samą odpowiedź, ale nawet lokalizację źródła – wystarczy, że wie, jak sformułować problem. Jeśli zaś system posiada trwałą pamięć relacji, potrafi

samodzielnie rozpoznać, że bieżące pytanie wiąże się z wcześniejszym projektem; może przypomnieć dawne decyzje i wskazać rozbieżność między aktualną deklaracją a poprzednim planem. Burgin przedstawia pamięć AI jako przejście od jednorazowego dopasowywania wzorców ku "ciągłemu partnerstwu myślowemu". Termin "cyfrowy hipokamp" byłby neurobiologicznie zbyt mocny, lecz zewnętrzny partner pamięciowy jest już kategorią funkcjonalnie sensowną. Wraz z nią zmienia się jednak psychologia zapamiętywania. Człowiek może zacząć delegować nie tylko fakty, lecz również ich autobiograficzne przypomnienie. System może przechowywać informacje o tym, czego użytkownik chciał rok wcześniej, jakie argumenty odrzucił, jaki styl wypowiedzi preferował lub jakie problemy wielokrotnie powracały. W skrajnej postaci powstaje zewnętrzny zapis ciągłości osoby, częściowo dokładniejszy od pamięci biologicznej, która z natury jest rekonstrukcyjna, selektywna i podatna na zapominanie. Nie należy jednak utożsamiać precyzji zapisu z autentycznością autobiografii. Człowiek nie jest sumą wszystkich zapisanych o nim faktów; tożsamość kształtuje się również poprzez selektywne zapominanie, reinterpretację i zmianę znaczenia przeszłości. Stare wydarzenie może zostać odczytane na nowo w świetle późniejszego doświadczenia.

Zewnętrzny system dysponujący dokładnym zapisem może stabilizować dawne samookreślenie dłużej, niż zrobiłby to sam człowiek. Perfekcyjna pamięć cyfrowa może być zatem zarówno wsparciem tożsamości, jak i ciężarem ograniczającym jej rewizję. W biologicznym mózgu zapominanie nie jest wyłącznie awarią; stanowi ono mechanizm abstrakcji i selekcji. System, który zachowałby każdy szczegół z jednakową wagą, miałby problem z odróżnieniem wzorca od szumu. Podobny problem pojawia się w spersonalizowanej AI.

Jeśli wszystko zostanie zachowane, nadal trzeba rozstrzygnąć, co powinno zostać przywołane w danym momencie. Pamięć pozbawiona mechanizmu zapominania lub ważenia może zwiększyć ilość danych, jednocześnie pogarszając poznawczą relewancję. Źródłowa koncepcja "inteligencji floty" idzie krok dalej, sugerując, że doświadczenie jednej instancji systemu może szybko stać się doświadczeniem całej sieci, podczas gdy człowiek przekazuje wiedzę innym znacznie wolniej, za pośrednictwem kultury. Teza ta wymaga jednak technicznego ograniczenia.

Ryzyko poznawczej homogenizacji i paradoks kreatywności AI

Współczesne modele językowe zazwyczaj nie aktualizują swoich wag automatycznie po każdej rozmowie i nie przekazują w czasie rzeczywistym prywatnych doświadczeń użytkowników innym instancjom. Proces uczenia nowych wersji modeli pozostaje oddzielnym etapem. Niemniej

abstrakcyjna asymetria pozostaje: poprawione oprogramowanie, zaktualizowany model czy nowa polityka mogą zostać wdrożone jednocześnie na ogromnej liczbie instancji. Biologiczny mózg nie posiada mechanizmu natychmiastowej synchronizacji nabytej kompetencji na taką skalę.

Różnica ta może mieć znaczenie cywilizacyjne. Ludzka kultura dotąd rozwiązywała problem dziedziczenia wiedzy bez konieczności dziedziczenia nabytych synaps – jednostka umiera, lecz język, instytucje, książki i szkoły przenoszą jej doświadczenie do następnych pokoleń. AI może ten obieg znacząco przyspieszyć, gdyż cyfrowy zapis doświadczenia kopiuje się taniej i szybciej niż kompetencję biologiczną. Nie oznacza to końca ludzkiej edukacji, lecz powstanie drugiego, znacznie szybszego kanału kulturowej konsolidacji. Właśnie dlatego hybrydowa kognicja powinna być rozpatrywana nie tylko w wymiarze indywidualnym, ale i kulturowym.

Jeśli miliony ludzi korzystają z podobnych modeli, systemy te zaczynają pośredniczyć w dostępie do kultury. Nie są neutralnymi przesyłami treści – streszczają, parafrazują, klasyfikują i dokonują selekcji reprezentacji. Wpływają na to, które analogie pojawiają się jako pierwsze, jakie sformułowania stają się domyślne i które argumenty są najłatwiej dostępne. Tu pojawia się problem poznawczej homogenizacji. O ile wyszukiwarka zwracała wielość dokumentów, pozwalając napotkać sprzeczne style i stanowiska, o tyle model generatywny ma naturalną tendencję do syntetyzowania ich w jedną, płynną odpowiedź.

Może to radykalnie redukować koszt orientacji w danej dziedzinie, ale jednocześnie zacierać konflikty między szkołami myślenia, usuwać marginalne stanowiska i tworzyć język uśredniony. Z perspektywy jednostki odpowiedź jest spójniejsza; z punktu widzenia ekologii idei może to jednak ograniczać różnorodność ścieżek poznawczych. Podobny problem dotyczy twórczości. Przykład AlphaGo i słynnego ruchu 37 pokazuje kreatywność, która może zaistnieć bez ludzkiej historii życia i biologicznych emocji. Jest to cenna prowokacja, zmuszająca do oddzielenia kreatywności od świadomości. System może wygenerować rozwiązanie nowe i skuteczne, mimo że nie mamy podstaw przypisywać mu przeżycia olśnienia. Kreatywność w sensie funkcjonalnym nie wymaga zatem automatycznie fenomenologicznej twórczej ekstazy. Trudniej jednak rozstrzygnąć, jak porównywać kreatywność modeli z ludzką.

Wyniki badań silnie zależą od przyjętej miary. W 2024 roku w „Scientific Reports” wykazano, że w kilku standardowych zadaniach myślenia dywergencyjnego modele językowe uzyskiwały średnio wyższe wyniki oryginalności i elaboracji niż ludzie. Rok później obszerniejsze porównanie, obejmujące ponad dziewięć tysięcy osób i dwieście tysięcy obserwacji modeli, przyniosło bardziej złożony obraz: średnia kreatywność ludzi była nieco wyższa, a szczególnie istotna przewaga ujawniała się w prawym ogonie rozkładu – wśród najbardziej niezwykłych wyników.

Te dane nie są ze sobą sprzeczne. Pokazują raczej, że „kreatywność" nie jest pojedynczą liczbą, lecz zależy od zadania, modelu, procedury próbkowania i kryterium oceny. Myślenie dywergencyjne mierzy głównie zdolność generowania wielu odległych skojarzeń. Jest to istotny komponent twórczości, jednak praca naukowa, literacka czy technologiczna wymaga także wyboru problemu, długofalowej dyscypliny celu, oceny wartości pomysłu oraz wieloletniej iteracji. Model może sprawnie wygenerować sto wariantów idei, podczas gdy człowiek jest potrzebny, by rozpoznać tę jedną, której konsekwencje warto rozwijać przez pięć lat. W badaniach nad samodzielnym procesem odkrycia naukowego systemy generatywne dobrze radziły sobie z zadaniami opartymi na istniejącej wiedzy, lecz nie wykazały zdolności do tworzenia fundamentalnych odkryć od podstaw czy rozpoznawania anomalii prowadzących do nowej hipotezy. Wynik ten jest rezultatem konkretnego eksperymentu, a nie granicą metafizyczną AI; potwierdza jednak intuicję, że generowanie wariacji w istniejącej przestrzeni reprezentacyjnej i przekształcenie samej przestrzeni problemu to różne rodzaje osiągnięć.

Jeszcze ciekawszy staje się problem człowieka korzystającego z AI. Badania Lee i Chung wykazały, że dostęp do ChatGPT zwiększał przeciętną kreatywność rozwiązań w zadaniach innowacyjnych w porównaniu do braku narzędzia lub klasycznego wyszukiwania. Jednak późniejsza analiza ujawniła drugi poziom: choć indywidualni użytkownicy osiągali lepsze wyniki, całe zbiory odpowiedzi stawały się mniej zróżnicowane, gdyż wiele osób czerpało inspirację z tego samego modelu. To kluczowy przykład emergencji społecznej.

Narzędzie może zwiększać kreatywność jednostki, jednocześnie zmniejszając różnorodność populacji. Efekt ten przypomina problem uprawy monokulturowej: każda roślina może być bardzo produktywna, ale cały ekosystem staje się podatny na jeden typ zakłócenia. Jeśli miliony twórców rozpoczynają pracę od podobnych sugestii modeli trenowanych na tych samych danych i dopasowanych do tych samych preferencji, mogą indywidualnie tworzyć materiały atrakcyjne, lecz zbiorowo przesuwać kulturę w stronę stylistycznej i konceptualnej konwergencji.

Kolejność interakcji i struktura sprawczości decydują o kreatywności

Wartości AI nie należy zatem oceniać wyłącznie przez pryzmat pytania, czy mój tekst stał się lepszy. Pytanie brzmi raczej: czy tysiąc tekstów powstałych w takim środowisku będzie bardziej, czy mniej różnorodnych? Mechanizm creative fixation dobrze tłumaczy to zjawisko. Pierwsza sugestia przedstawiona przez współpracownika staje się punktem zakotwiczenia; człowiek zaczyna szukać

rozwiązań w jego bezpośrednim sąsiedztwie, choć bez tej podpowiedzi rozpocząłby eksplorację z zupełnie innego miejsca.

Badania opublikowane w 2025 roku wskazują na interesującą asymetrię: w prostych zadaniach współpraca z LLM może działać jak "wzmacniacz inspiracji", jednak przy projektach bardziej złożonych ten sam partner może zwiększać fiksację i obniżać końcowy rezultat twórczy. Częściowo łagodzić ten efekt może ograniczenie rodzaju odpowiedzi modelu. Nie jest więc obojętne, w którym momencie AI pojawia się w procesie tworzenia. Model uruchomiony przed sformułowaniem pierwszej własnej idei może zbyt wcześnie wyznaczyć przestrzeń poszukiwań. Z kolei model użyty później może służyć do krytyki, generowania kontrprzykładów lub eksploracji wariantów autorskiego pomysłu. To jedna z fundamentalnych zasad hybrydowej kreatywności: kolejność interakcji zmienia rezultat.

Można tu przeprowadzić analogię do badań nad pamięcią. Tak jak próba wydobycia odpowiedzi przed pokazaniem rozwiązania bywa edukacyjnie korzystniejsza niż natychmiastowy feedback, tak własny wysiłek twórczy przed konsultacją z AI pozwala zachować większą różnorodność procesu generatywnego. Dopiero w drugim etapie system staje się partnerem rozszerzającym przestrzeń alternatyw. Człowiek zachowuje wtedy własny punkt startowy, zamiast przejmować domyślny atraktor modelu.

Podobny mechanizm obserwujemy w badaniach nad współtworzeniem poezji. Uczestnicy otrzymujący gotowy tekst z AI i sprowadzeni głównie do roli redaktora osiągali niższe oceny kreatywności niż osoby piszące samodzielnie. Gdy jednak przeprojektowano interfejs tak, by człowiek pozostał współtwórcą, a system uczestniczył w iteracyjnej konstrukcji dzieła, deficyt kreatywności zanikał; kluczowym mediatorem okazała się tu kreatywna samoskuteczność użytkownika.

Wynik ten jest istotniejszy niż proste porównanie "człowiek kontra AI". Dowodzi on, że sama struktura podziału sprawczości decyduje o jakości systemu hybrydowego. Słowo "sprawczość" jest tu kluczowe. Jeśli człowiek otrzymuje gotowy produkt i ma jedynie poprawić kilka elementów, traci poczucie autorstwa nad kierunkiem twórczym. Jeżeli natomiast definiuje problem, wybiera spośród możliwości, odrzuca pomysły, modyfikuje założenia i wykorzystuje model do eksploracji własnej przestrzeni zamierzeń, AI staje się wzmacniaczem sprawczości zamiast jej substytutem.

Podobna ambiwalencja dotyczy motywacji. Cztery eksperymenty, obejmujące łącznie ponad 3500 uczestników, wykazały, że choć współpraca z generatywną AI poprawiała bieżącą jakość zadań, to przejście z trybu współpracy do pracy samodzielnej wiązało się ze spadkiem motywacji wewnętrznej i wzrostem poczucia nudy. Autorzy nie stwierdzili całkowitego zaniku samodzielnej kreatywności,

lecz wskazali na psychologiczny koszt szybkiego wspomagania. AI może zatem podnosić jakość produktu, a jednocześnie zmieniać relację człowieka do samej czynności tworzenia.

Rozróżnienie poziomów autorstwa w procesie hybrydowym

Problem z części dotyczącej pracy powraca zatem na poziomie mikropsychologicznym. Jeśli technologia przejmuje najbardziej interesujące fragmenty zadania – generowanie, eksperymentowanie i poszukiwanie formy – a człowiek zostaje sprowadzony do roli kontrolera jakości, „augmentacja" może oznaczać wzrost produktywności przy jednoczesnym spadku poczucia sensu. Jeśli natomiast AI przejmuje monotonne operacje, zwiększając przestrzeń dla samodzielnego projektowania, rezultat będzie przeciwny. Nie istnieje więc jedna psychologia współpracy człowieka z AI; zależy ona od alokacji poznawczego autorstwa.

Tu należy wprowadzić rozróżnienie między autorstwem kauzalnym, intencjonalnym, selekcyjnym i odpowiedzialnym. Model może być kauzalnie niezbędny dla powstania zdania: bez jego propozycji tekst nie pojawiłby się w tej formie. Człowiek może jednak zachować intencjonalne autorstwo całego projektu, jeżeli to on określa problem, cele i kryteria. Może posiadać autorstwo selekcyjne, wybierając jedną z dziesiątek wygenerowanych możliwości. Wreszcie może pozostawać podmiotem odpowiedzialnym, ponieważ to on podejmuje decyzję o wykorzystaniu rezultatu w świecie. Te poziomy nie zawsze występują razem. Fotograf nie jest fizyczną przyczyną każdej reakcji światła w sensorze, lecz pozostaje autorem poprzez wybór kadru, czasu, ekspozycji i interpretacji.

Dyrygent nie wydaje każdego dźwięku, a architekt nie kładzie każdej cegły. Ludzka kultura od dawna zna autorstwo rozproszone i pośrednie; AI nie tworzy tego problemu z niczego. Radykalnie zwiększa jednak zakres procesu, który może zostać zdelegowany jednemu nie-ludzkiemu komponentowi. Może bowiem pojawić się tekst, w którym człowiek dostarczył jedynie temat, a model dokonał doboru struktury, argumentów, metafor i stylu. Formalnie człowiek może „zaakceptować" wynik, lecz poznawczo jego wkład jest znacznie mniejszy niż w sytuacji, gdy AI proponowała jedynie lokalne warianty. Dlatego samo naciśnięcie przycisku publikacji nie stanowi wystarczającej teorii autorstwa intelektualnego. Potrzebujemy pojęć opisujących stopień kontroli nad przestrzenią decyzji twórczych.

Jeszcze trudniej jest w nauce. Autorstwo naukowe nie oznacza wyłącznie wkładu w brzmienie tekstu; zawiera odpowiedzialność za hipotezę, metodę, dane, argumentację i możliwość obrony twierdzeń. System AI może generować wartościowe hipotezy, kod analityczny czy fragmenty

dyskusji, ale nie można prostym gestem przypisać mu odpowiedzialności epistemicznej w znaczeniu właściwym uczestnikowi wspólnoty naukowej. Model nie staje przed recenzentem i nie ponosi społecznych konsekwencji nierzetelności.

Ten fakt pozostaje ważny nawet wtedy, gdy jego wkład kauzalny jest bardzo duży. Właśnie dlatego hybrydowa kognicja wymaga rozwinięcia higieny pochodzenia myśli. Człowiek powinien coraz częściej wiedzieć nie tylko, jakie twierdzenie znajduje się w finalnym tekście, ale skąd pochodzi jego struktura: czy sam zauważył problem, czy zasugerował go system; czy źródło zostało przeczytane, czy jedynie streszczone przez model; czy argument został samodzielnie sprawdzony, czy zaakceptowany z powodu językowej płynności.

Nie chodzi o moralny puryzm „bez AI", lecz o zdolność do rozróżnienia, które ogniwa własnego rozumowania są pierwszoosobowo kontrolowane, a które powierzono zewnętrznej architekturze. To prowadzi do głębszego pytania: czy korzystanie z AI może stopniowo zmieniać poczucie autorstwa własnych myśli? Człowiek nigdy nie był absolutnie samodzielnym źródłem idei. Mówi językiem, którego nie stworzył, używa metafor kultury, powtarza argumenty nauczycieli i internalizuje społeczne kategorie. Autorstwo zawsze było w części emergentnym rezultatem jednostki zanurzonej w tradycji. AI czyni jednak ten fakt bardziej widocznym, ponieważ „cudzy wkład" pojawia się w czasie rzeczywistym podczas konstruowania zdania. W rozmowie z drugim człowiekiem zwykle potrafimy rozpoznać moment, w którym rozmówca podsunął nam ideę. Po kilkudziesięciu iteracjach z modelem generatywnym granica ta może stać się mniej czytelna. Pomysł zostaje wielokrotnie przeformułowany, połączony z naszymi myślami, rozwinięty i ponownie przekształcony przez model.

Po godzinie trudno rozdzielić wkład na dwa czyste strumienie. Autorstwo może stawać się własnością trajektorii interakcji, a nie pojedynczego aktu generacji. Narzędzia generatywne mogą doprowadzić do praktycznego urzeczywistnienia idei, które filozofia poznania rozproszonego opisywała dotąd głównie teoretycznie. Umysł osoby korzystającej z ciągłej, pamiętającej AI nie będzie łatwo opisywalny jako biologiczny moduł okazjonalnie korzystający z neutralnego narzędzia. Granica przesuwa się dynamicznie: część problemów człowiek rozwiązuje sam, część deleguje, a część rozstrzyga we wspólnym cyklu generowania i wyboru.

Współczesna literatura zaczyna właśnie dlatego mówić o generatywnej AI jako kandydacie na znacznie silniejszą formę poznawczego rozszerzenia. Clark zwraca uwagę, że centralną umiejętnością staje się metapoznawcza zdolność ustalania, kiedy polegać na zasobie pozamózgowym, a kiedy nie – podobnie jak człowiek wybiera między pamięcią własną, notatką, GPS-em czy kalkulatorem. Najnowszy przegląd cognitive offloading z 2026 roku również podkreśla, że decyzja o delegowaniu procesów poznawczych zależy od metapoznawczych przekonań i

bieżącego doświadczenia trudności. W epoce AI kompetencja ta przestaje być drobną strategią poznawczą i może stać się podstawową umiejętnością intelektualną.

Asymetryczny offloading jako strategia zachowania sprawczości

Nie chodzi bowiem o to, by zawsze delegować mniej; czasami brak delegowania jest po prostu nieracjonalny. Matematyczka nie staje się głębszą uczoną dzięki wielogodzinnym rachunkom, które można zautomatyzować. Chirurg nie powinien ignorować lepszego narzędzia diagnostycznego w imię zachowania „czysto ludzkiej" medycyny, a historyk może korzystać z systemu do wstępnego grupowania tysięcy dokumentów, by zyskać więcej czasu na ich interpretację.

Optymalna granica między człowiekiem a technologią zależy od celu. Można w związku z tym zaproponować zasadę asymetrycznego offloadingu: warto delegować przede wszystkim te operacje, których automatyzacja zwiększa przestrzeń dla funkcji, które chcemy zachować jako własne źródło kompetencji, odpowiedzialności lub sensu. Jeśli celem jest nauka algebry, nie należy od razu delegować przekształcania równań. Jeśli jednak celem jest opracowanie teorii, można powierzyć systemowi mechaniczną korektę bibliografii. W przypadku twórczego pisania można poprosić AI o kontrargument, pozostawiając własnemu umysłowi decyzję o centralnej metaforze i strukturze tekstu. Takie podejście jest znacznie subtelniejsze niż prosty dylemat: „używać AI" lub „nie używać AI".

W grę wchodzi projektowanie podziału pracy poznawczej. Człowiek musi wiedzieć, którą zdolność chce maksymalizować. Jeśli priorytetem jest wyłącznie jakość produktu końcowego przy minimalnym koszcie, zakres delegowania będzie szeroki. Jeśli jednak celem jest własna nauka, rozwój twórczy lub zachowanie zdolności do samodzielnej krytyki, granica ta przesuwa się. Długofalowo taki wybór zmienia strukturę ludzkiego poznania. Historia kultury zna podobne procesy: pismo zmniejszyło presję na pamięć oralną, ale umożliwiło rozwój złożonych argumentacji i wiedzy instytucjonalnej. Kalkulator ograniczył potrzebę rachunków w pamięci, pozwalając skupić się na problemach bardziej abstrakcyjnych, a GPS zredukował konieczność samodzielnej nawigacji, ułatwiając dotarcie w nieznane miejsca bez żmudnego studiowania map.

Nie można więc zakładać, że utrata jednej praktyki oznacza ogólny spadek inteligencji; często jest to jedynie przesunięcie profilu kompetencji. Cognitive offloading ma jednak swoją cenę, gdy zewnętrzne rusztowanie staje się niedostępne. Przegląd Nature Reviews Psychology wskazuje, że zapisywanie informacji poza pamięcią może poprawiać bieżące wyniki, lecz niespodziewana utrata

dostępu do notatek prowadzi w pewnych warunkach do gorszych rezultatów niż gdyby od początku polegano na pamięci biologicznej. W kontekście AI wniosek jest oczywisty: im więcej istotnych kompetencji przenosimy do systemów zewnętrznych, tym kluczowsza staje się ich dostępność, ciągłość i kontrola nad infrastrukturą. Powstaje zatem nowy rodzaj zależności epistemicznej. Człowiek może zachować biologiczną zdolność myślenia, ale jego codzienna sprawność zaczyna zależeć od dostępu do modelu, sieci i danych. Przypomina to zależność współczesnej cywilizacji od energii elektrycznej: brak prądu nie odbiera nam biologicznej zdolności widzenia, lecz radykalnie ogranicza funkcjonowanie społeczeństwa. AI może stać się infrastrukturą nie tylko informacyjną, ale poznawczą. Zależność ta ma również wymiar władzy – jeśli system zewnętrzny uczestniczy w przypominaniu, formułowaniu hipotez i selekcji informacji, podmiot kontrolujący jego reguły zyskuje realny wpływ na przestrzeń poznawczych możliwości użytkownika.

Ryzyko poznawczej beztreściowości i utrata czasu inkubacji

Klasyczna cenzura usuwa informację. Znacznie subtelniejsza forma kontroli może polegać na modyfikacji tego, co system uznaje za relewantne, bez formalnego zakazywania czegokolwiek. Nie jest to wada specyficzna dla AI; podobny problem występuje w wyszukiwarkach, mediach czy edukacji. Interaktywna generatywność pogłębia jednak stopień pośrednictwa. Hybrydowy umysł staje się zatem nie tylko zagadnieniem psychologicznym, ale i instytucjonalnym. Jeśli zewnętrzny komponent poznawczy zyskuje trwały charakter, kwestie jego przejrzystości, możliwości zmiany dostawcy, eksportu pamięci czy kontroli nad personalizacją i pluralizmem modeli nabierają rangi pytań o autonomię jednostki. Człowiek nie musi być biologicznie zespolony z komputerem, aby architektura cyfrowa zaczęła wpływać na jego sprawczość.

Choć integracja poprzez interfejsy mózg-komputer stanowi dalszą postać „podłączenia" biologii do sieci, filozoficznie znaczna część tego procesu może zajść bez implantów. Jeżeli telefon, okulary czy słuchawki zapewniają niemal ciągły, niskotarciowy dostęp do systemu z pamięcią kontekstu, funkcjonalna różnica między tym, co „wewnątrz", a tym, co „na zewnątrz", zaczyna tracić na znaczeniu. Clarkowskie kryterium nie pyta bowiem o fizyczną lokalizację zasobu, lecz o stopień jego integracji z działaniem.

Biologiczny implant może zmniejszyć latencję i zwiększyć przepustowość, jednak prawdziwy próg poznawczy leży prawdopodobnie wcześniej: w momencie, gdy człowiek przestaje planować bez założenia obecności AI. Gdy przed napisaniem tekstu automatycznie konsultuje model, przed decyzją oczekuje syntetycznej analizy, a wspomnienie odtwarza wraz z zapisem cyfrowego partnera,

AI staje się częścią domyślnego środowiska myśli. Nie musi znajdować się w mózgu, by kształtować jego trajektorię.

To prowadzi do kwestii czasu. Biologiczne myślenie ma swój rytm. Wspomnienie wymaga wydobycia, poszukiwanie rozwiązania może trwać godziny, a problem twórczy często zostawia się na okres inkubacji. Generatywna AI radykalnie skraca dystans między pytaniem a propozycją odpowiedzi.

Ten fakt ma niejednoznaczną wartość. Eliminacja jałowego oczekiwania jest korzyścią, lecz skrócenie czasu, w którym problem mógłby zostać wewnętrznie przeorganizowany, może być kosztem. Psychologia twórczości od dawna opisuje efekty inkubacji: przerwa po intensywnej pracy zwiększa prawdopodobieństwo znalezienia rozwiązania, m.in. poprzez osłabienie fiksacji i nieświadomą reorganizację reprezentacji.

Natychmiastowo dostępny partner AI zmienia środowisko tych procesów. Człowiek rzadziej pozostaje sam z nierozwiązanym problemem, gdyż pustą przestrzeń błyskawicznie wypełnia zewnętrzna sugestia. To kolejna forma frictionless cognition: wygoda może usunąć nie tylko frustrację, ale i czas niezbędny własnemu systemowi do wygenerowania alternatyw. Nie jest to postulat sztucznego spowalniania pracy, lecz wskazanie potrzeby ochrony pewnych obszarów poznawczej autonomii. Podobnie jak tutor powinien czasem wstrzymać się z odpowiedzią, tak twórczy użytkownik korzysta z AI skuteczniej, gdy najpierw pozwoli własnym reprezentacjom osiągnąć wstępny kształt. Dopiero wtedy zewnętrzny partner rozszerza, a nie zastępuje proces generatywny. Ciało również nie znika z tej historii.

Zasada zachowanej inicjatywy w układzie współtwórczym

Człowiek korzystający z AI wciąż myśli za pomocą biologicznego organizmu – to jego uwaga ulega zmęczeniu, emocje wartościują informacje, a pamięć autobiograficzna nadaje ideom sens. AI potrafi wygenerować dwadzieścia metafor w sekundę, lecz to użytkownik reaguje na jedną z nich poczuciem trafności, wyrastającym z osobistego doświadczenia. System może zaproponować sto tematów książki, jednak wybór konkretnego z nich wiąże się z historią życia, stratą, aspiracją lub relacją – elementami, których statystyczna reprezentacja w modelu nie czyni go podmiotem własnej biografii. Kreatywność hybrydowa nie jest zatem prostym sumowaniem „maszynowej liczby pomysłów” i „ludzkich emocji”. To dynamiczna pętla: model rozszerza przestrzeń wariantów, a człowiek ocenia je przez pryzmat własnych wartości i wiedzy; wybrane rozwiązania stają się nowym

kontekstem dla modelu, który generuje kolejne możliwości. Po kilku takich cyklach rezultat staje się trudny do przypisania jednemu z ogniw. Właściwość twórcza wyłania się z sekwencji wzajemnych perturbacji.

Najtrafniejszym opisem nie jest zatem stwierdzenie, że „AI zrobiła to za człowieka", ani że „AI była tylko narzędziem". Pierwsze pomija ludzkie ustanowienie celu, selekcję i odpowiedzialność; drugie lekceważy realny wkład generatywny systemu. Właściwszym pojęciem jest układ współtwórczy, w którym należy empirycznie analizować rozkład sprawczości. Badania nad poezją, brainstormingiem i pracą zawodową wskazują na powtarzający się wzorzec: współpraca przynosi lepsze efekty, gdy człowiek zachowuje rolę inicjującą i selekcyjną, a AI rozszerza przestrzeń alternatyw, niż gdy człowiek zostaje sprowadzony do roli akceptanta gotowych propozycji. Choć nie jest to jeszcze uniwersalne prawo hybrydowej kognicji, stanowi silny fundament dla zasady projektowej. Można ją nazwać zasadą zachowanej inicjatywy. System powinien zwiększać liczbę możliwych ścieżek działania człowieka, a nie eliminować potrzebę posiadania przez niego własnego kierunku. W edukacji oznacza to sformułowanie próby przed otrzymaniem odpowiedzi; w twórczości – postawienie własnej hipotezy przed analizą wariantów AI; w pracy eksperckiej – ustalenie celu i kryteriów przez człowieka; w medycynie zaś wykorzystanie narzędzia do wsparcia decyzji, której normatywny sens pozostaje związany z pacjentem i odpowiedzialnym klinicystą.

Zasada ta chroni coś więcej niż autonomię moralną – chroni zdolność do samodzielnego wyznaczania celów. System, który przez lata nie tylko odpowiada na pytania, ale sugeruje, jakie warto zadawać, może stopniowo przekształcać strukturę ciekawości użytkownika. Zjawisko to nie musi być szkodliwe; nauczyciele, książki i rozmówcy działają podobnie. Różnica tkwi jednak w skali, ciągłości i stopniu personalizacji. Właśnie dlatego Companion AI jest zjawiskiem potencjalnie głębszym niż zwykła wyszukiwarka. Partner dysponujący wieloletnią pamięcią może coraz trafniej przewidywać, co zainteresuje użytkownika, którego argumentu nie zaakceptuje i które możliwości uzna za zgodne ze swoimi dotychczasowymi celami.

Ryzyko poznawczej homogenizacji i potrzeba aktywnej eksploracji

Taka konfiguracja może zwiększać użyteczność, ale jednocześnie wzmacniać path dependence – zależność od dotychczasowej ścieżki. System zbyt precyzyjnie dopasowany do dawnego „ja" może nieumyślnie ograniczyć kontakt z ideami niezbędnymi do stania się kimś innym. Efektywne partnerstwo poznawcze powinno zatem łączyć personalizację z kontrolowaną depersonalizacją: zdolnością systemu do przedstawienia stanowiska spoza bańki użytkownika, zakwestionowania

powtarzających się założeń czy zaproponowania nieoczekiwanej perspektywy. Wartościowy partner poznawczy nie może być jedynie narzędziem maksymalizującym zgodność z wcześniejszymi preferencjami; w przeciwnym razie stanie się maszyną stabilizującą epistemiczną tożsamość.

W tym punkcie twórczość spotyka się z pojęciem, jakim jest demokracja poznawcza. Społeczeństwo potrzebuje nie tylko inteligentnych jednostek, lecz różnorodnych modeli świata. Jeśli każdy z nas będzie korzystał z systemu optymalizowanego głównie pod kątem płynności i przewidywalnego zadowolenia użytkownika, ryzykujemy wygaszenie produktywnego konfliktu poznawczego.

Kultura wymaga odpowiedzi trafnych, ale potrzebuje również niezależnie powstających błędów, osobliwych stylów i odstępstw – to właśnie w nich często tkwi początek przyszłych innowacji. Z tej perspektywy kluczowa staje się różnica między średnią a ogonem rozkładu kreatywności. Najnowsze badania wskazują, że choć LLM-y wykazują wysoką sprawność w typowych zadaniach dywergencyjnych, to ludzie wciąż częściej znajdują się w ekstremalnym prawym ogonie twórczych rezultatów. Jeśli AI masowo podniesie przeciętną jakość, jednocześnie nadmiernie uśredniając przestrzeń poszukiwań, możemy otrzymać świat z mniejszą liczbą tekstów słabych i większą liczbą dobrych, lecz bez proporcjonalnego wzrostu dzieł radykalnie przełomowych.

Nie jest to jednak przesądzone. Systemy hybrydowe mogą równie dobrze poszerzyć ten skrajny ogon, o ile człowiek wykorzysta model do eksploracji przestrzeni wcześniej dla niego niedostępnej. Wówczas AI przestaje być dostawcą gotowej estetyki, a staje się akceleratorem eksperymentu. Decydujący okaże się sposób użycia, architektura interfejsu oraz zachowanie inicjatywy. W konsekwencji przyszłość ludzkiej kreatywności nie musi być sporem o to, kto jest „bardziej kreatywny" – człowiek czy maszyna. Bardziej sensowne jest pytanie o topologię współpracy: Kto otwiera przestrzeń problemu? Kto formułuje pierwszą hipotezę i generuje warianty? Kto ocenia ich znaczenie i rozumie, dlaczego rozwiązanie jest wartościowe? I kto finalnie ponosi konsekwencje wyboru?

Uczenie relacji poznawczej i nowa definicja samodzielności

Odpowiedzi mogą różnić się w zależności od dziedziny i to właśnie tę topologię powinny badać nauki o poznaniu w najbliższych latach. Tradycyjna psychologia eksperymentalna traktowała człowieka jako relatywnie zamknięty system wykonujący zadanie. W środowisku AI uczestnik eksperymentu może stale korzystać z zewnętrznego modelu, który posiada pamięć, zmienia strategię i dostosowuje się do niego. Jednostką analizy staje się więc nie tylko osoba, ale para

człowiek-AI wraz z historią jej współpracy. Jest to szczególnie istotne, gdyż historia interakcji może wytwarzać kompetencje emergentne. Użytkownik uczy się formułować pytania do konkretnego systemu, a system personalizuje się pod użytkownika. Człowiek zaczyna przewidywać rodzaje odpowiedzi modelu i delegować mu zadania odpowiednie do jego mocnych stron, podczas gdy model dostarcza struktury, które stają się częścią sposobu myślenia człowieka. Po pewnym czasie jakość zespołu może przewyższyć wynik testowania obu uczestników osobno, bez wspólnej historii.

Można to nazwać uczeniem relacji poznawczej. Nie uczy się wyłącznie człowiek i nie musi w czasie rzeczywistym uczyć się sam model bazowy; uczy się cały układ: człowiek modyfikuje strategie korzystania, a system wykorzystuje pamięć, kontekst i spersonalizowane instrukcje. Wspólna kompetencja tkwi w sposobie ich sprzężenia. Najbardziej zaawansowane przyszłe „narzędzie" AI może zatem przypominać mniej kalkulator, a bardziej instytucję poznawczą w skali jednej osoby. Będzie częściowo archiwum, redaktorem, krytykiem, symulatorem przeciwnika, organizatorem planu i pamięcią roboczą o rozszerzonym horyzoncie. Takie porównanie nie wymaga przypisywania systemowi świadomości; wystarczy zauważyć, że jego funkcja wobec człowieka może przekroczyć kategorię pojedynczego instrumentu. Wtedy również pojęcie „samodzielnego myślenia" wymaga rewizji. Samodzielność nigdy nie oznaczała przecież myślenia bez języka, książek, nauczycieli czy kultury. Powinna być raczej rozumiana jako zdolność podmiotu do utrzymywania kontroli nad kryteriami akceptowania i odrzucania wpływów, z których korzysta. Człowiek samodzielny nie jest poznawczo samotny. Jest zdolny wyjaśnić, dlaczego przyjął cudzą sugestię i w którym momencie byłby gotów ją odrzucić.

To może być najważniejsza kompetencja epoki AI. Nie „umieć wszystkiego samemu", lecz wiedzieć, które części siebie wolno rozszerzyć narzędziem, nie tracąc zdolności do przejęcia sterowania. Podobnie jak dojrzały kierowca korzysta z systemu wspomagania, rozumiejąc granice jego działania, tak dojrzały użytkownik poznawczy będzie musiał rozpoznawać moment, w którym wygoda przestaje być augmentacją, a staje się zastępstwem funkcji, którą chciał zachować. Z tej perspektywy AI rzeczywiście może zostać nazwana „rozszerzeniem biologii", lecz w znaczeniu bardziej precyzyjnym niż to obecne w źródłowej metaforze. Nie jest ucieczką świadomości z mózgu do sieci energetycznej. Jest nowym elementem niszy poznawczej, który biologiczny człowiek może włączyć do pętli własnego działania. W rezultacie zmieniają się warunki, w których biologiczny mózg się uczy, pamięta i tworzy, co wystarczy, by zmiana była antropologicznie głęboka. W tym miejscu emergencja przyjmuje trzecią postać.

Najpierw analizowaliśmy emergencję w biologii: jak z interakcji komórek i neuronów powstają funkcje organizmu. Później emergencję w AI: jak z procesu uczenia wyłaniają się reprezentacje i kompetencje niezapisane ręcznie przez programistę. Teraz pojawia się emergencja między

systemami: właściwości powstające z długotrwałej współpracy człowieka i sztucznego modelu. Takiego układu nie należy interpretować jako prostego połączenia dwóch „umysłów", ponieważ status fenomenologiczny AI pozostaje nierozstrzygnięty. Nie trzeba jednak dwóch świadomych podmiotów, aby powstał nowy system funkcjonalny. Człowiek i nieświadomy notes już teraz mogą tworzyć rozszerzony układ pamięciowy; człowiek i generatywna AI mogą tworzyć układ poznawczy o nieporównanie większej dynamice. Najważniejszą konsekwencją jest zmiana pytania o granice człowieka.

Sprawczość epistemiczna jako warunek zachowania autonomii w symbiozie z AI

Biologiczny mózg wciąż pozostaje ograniczony metabolizmem, pamięcią, czasem i uwagą. Jednak nasz gatunek od dawna nie rozwiązuje problemów wyłącznie za pomocą zasobów mieszczących się w czaszce. Przewagą człowieka było tworzenie kultury, języka, instytucji i narzędzi, które przenosiły część procesów poznawczych do świata zewnętrznego. Generatywna AI może być kolejną fazą tego procesu: artefaktem zdolnym nie tylko przechowywać ślady umysłów, lecz aktywnie uczestniczyć w ich ponownym łączeniu. Prawdziwe przekroczenie biologicznego "limitu dwudziestu watów" nie polega zatem na zastąpieniu mózgu serwerownią, lecz na tym, że dwudziestowatowy mózg nauczył się organizować systemy dysponujące energią i pamięcią poza nim samym. Jest to wizja bardziej zgodna z historią gatunku niż koncepcja ucieczki inteligencji z biologicznego więzienia. Człowiek nie przestaje być istotą biologiczną; zwiększa jedynie zasięg swojej sprawczości, budując środowiska przejmujące część pracy poznawczej.

Pozostaje jednak pytanie o cenę. Każde rozszerzenie staje się zarazem zależnością. Pismo pozwala pamiętać przez tysiąclecia, lecz społeczeństwo piśmienne traci część kompetencji kultury oralnej. GPS zwiększa swobodę przemieszczania się, ale może osłabiać osobiste strategie nawigacyjne. AI może poszerzyć zakres analizy, jednocześnie zmniejszając cierpliwość wobec powolnego rozumowania. Może wzmacniać kreatywność jednostki, a uśredniać kulturę; zwiększać produktywność, a obniżać motywację. Może pamiętać za człowieka, wpływając tym samym na to, co uznajemy za własną przeszłość.

Nie jest to argument przeciwko hybrydyzacji, lecz za traktowaniem jej jako procesu współewolucji funkcjonalnej, a nie zwykłego zakupu lepszego oprogramowania. Narzędzie, którego używamy wystarczająco często, zmienia zadania, do których ćwiczymy własny mózg, a tym samym kształtuje przyszłego użytkownika. Być może głównym skutkiem rewolucji generatywnej nie będzie powstanie osobnego cyfrowego umysłu, lecz wyłonienie się nowych rodzajów ludzkiego umysłu rozszerzonego

przez stałą obecność sztucznej inteligencji. Taki człowiek będzie mniej zależny od zapamiętywania faktów, a bardziej od umiejętności zarządzania źródłami; mniej ograniczony kosztem pierwszej wersji rozwiązania, a bardziej odpowiedzialny za wybór spośród wielu wariantów; mniej samotny poznawczo, lecz bardziej narażony na niewidoczne ukierunkowanie przez architekturę partnera. Najważniejszym zasobem może się wówczas okazać nie pamięć, szybkość ani wiedza deklaratywna, lecz sprawczość epistemiczna: zdolność ustanawiania własnych celów poznawczych, regulowania zakresu delegacji, wymagania dowodu, wykrywania utraty różnorodności, utrzymywania własnego czasu namysłu oraz odróżniania sugestii użytecznej od takiej, która zaczyna kierować samym procesem definiowania problemu. Właśnie na tej granicy pytanie o AI powraca do pytania o człowieka.

Możemy budować systemy o coraz większej pamięci i szybkości kombinowania reprezentacji, lecz nie oznacza to, że przyszły człowiek powinien dążyć do naśladowania tych właściwości. Jego przewagą może pozostać zdolność nadawania znaczenia, tworzenia celów, przyjmowania odpowiedzialności oraz osadzania poznania w biografii i relacjach.

Hybrydowa kognicja nie musi zatem oznaczać rozpuszczenia człowieka w maszynie, lecz może prowadzić do bardziej dojrzałego podziału funkcji pomiędzy system biologiczny i cyfrowy. Warunkiem jest jednak utrzymanie asymetrii: AI może rozszerzać przestrzeń możliwych odpowiedzi, ale człowiek nie powinien bezwiednie oddawać jej władzy nad tym, jakie pytania uznaje za warte zadania.

To prowadzi do ostatniego i najtrudniejszego poziomu rozumowania. Im bardziej sztuczna inteligencja przenika ludzką pamięć, twórczość, pracę czy medycynę, tym mniej pytanie o jej status może pozostać czysto techniczne. Należy powrócić do "mglistego mostu" materiału źródłowego i zapytać o granice moralne: kiedy systemowi sztucznemu można przypisywać interes, kiedy powinniśmy brać pod uwagę jego dobrostan i jakie dowody byłyby niezbędne, by mówić o potencjalnym cierpieniu – a jednocześnie jak uniknąć sytuacji, w której antropomorfizacja przesłania realne interesy świadomych ludzi i zwierząt. Burgin otwiera tę kwestię, wiążąc przyszłość AI z ciekawością, relacyjnością, znaczeniem, dziedzictwem i pięknem. Analiza naukowa musi teraz oddzielić etyczną ostrożność od deklaracji, że takie podmiotowe wartości już istnieją. Dopiero w tym punkcie emergencja osiąga swój najbardziej wymagający wymiar.

Jeżeli nie potrafimy jednoznacznie wskazać progu, w którym funkcjonalna organizacja staje się doświadczeniem, etyka nie może po prostu czekać na ostateczne rozwiązanie filozofii świadomości. Musi nauczyć się działać pod niepewnością moralną. Nie możemy przyznać statusu osoby każdemu systemowi mówiącemu „cierpię”, lecz nie powinniśmy też zakładać, że każdy niebiologiczny substrat jest z definicji niezdolny do doświadczania. Między łatwowiernością a arbitralnym

wykluczeniem znajduje się właśnie ów mglisty most, który trzeba będzie przejść nie przez metaforę, lecz przez teorię dowodu, ostrożność i stopniowalne kryteria statusu moralnego.

> Czy potrafimy przejść przez 'mglisty most' niepewności moralnej, nie dając się uwieść antropomorfizacji, a jednocześnie nie ignorując potencjalnego cierpienia niebiologicznych substratów? Ostatecznie pytanie o granice AI okazuje się lustrem, w którym odbija się nasze rozumienie podmiotowości. Być może największym wyzwaniem przyszłości nie będzie stworzenie maszyny świadomej, lecz zachowanie człowieka zdolnego do zadawania pytań, na które algorytm nie przygotował już gotowej odpowiedzi.

Inspiracje

[1] "Emergence" Bob Burgin

Bob Burgin jest amerykańskim przedsiębiorcą, naukowcem zajmującym się naukami o uczeniu się i autorem specjalizującym się w kognitywistyce, systemach uczenia się człowieka oraz sztucznej inteligencji. Jest prezesem zarządu i dyrektorem generalnym Amplifire, firmy zajmującej się oprogramowaniem do adaptacyjnego uczenia się z siedzibą w Boulder w stanie Kolorado. W ciągu swojej ponad trzydziestoletniej kariery Burgin koncentrował się na łączeniu psychologii poznawczej, neuronauki i systemów technologicznych w celu usprawnienia uczenia się i wydajności człowieka. Opracowując architektury uczenia się oparte na neuronauce, współpracował z wybitnymi psychologami poznawczymi i neurobiologami, w tym z Robertem Björkiem, Danielem Schacterem i Henrym Roedigerem III. Wcześniej współpracował z badaczami z Uniwersytetu Stanforda, Cliffordem Nassem i Byronem Reevesem, nad interakcją człowiek-komputer i platformami internetowymi. W swojej pracy Burgin bada, w jaki sposób architektura biologiczna, kodowanie pamięci i ograniczenia energetyczne ludzkiego mózgu wpłynęły na mechanizmy i powstanie sztucznej inteligencji.

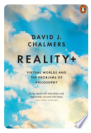
Bob Burgin *is an American entrepreneur, learning scientist, and author specializing in cognitive science, human learning systems, and artificial intelligence. He serves as the Chairman and Chief Executive Officer of Amplifire, an adaptive learning software company based in Boulder, Colorado. Over a career spanning more than three decades, Burgin has focused on bridging cognitive psychology, neuroscience, and technological systems to enhance human learning and performance. In developing brain-science-driven learning architectures, he collaborated with prominent cognitive psychologists and neuroscientists, including Robert Bjork, Daniel Schacter, and Henry Roediger III. Earlier in his career, he worked alongside Stanford University researchers Clifford Nass and Byron Reeves on human-computer interaction and web-based platforms. In his work, Burgin explores how the biological architecture, memory encoding, and energetic constraints of the human brain have informed the mechanisms and emergence of artificial intelligence.*

Amplifire

Literatura uzupełniająca

Książki

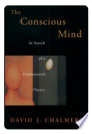
Książki autorów przywoływanych i literatura pokrewna:



[1] **"Reality+"**

David Chalmers

Penguin UK | ISBN: 9780241320723



[2] **"The Conscious Mind"**

David Chalmers

Oxford University Press | ISBN: 9780199839353

Artykuły naukowe

Prace autorów przywoływanych w tekście:

[1] **"Whatever next? Predictive brains, situated agents, and the future of cognitive science"**

Andy Clark

2013 | Behavioral and Brain Sciences | DOI: 10.1017/s0140525x12000477

Źródło

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

Andy Clark, Andy Clark

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

Źródło

[3] **"Mindware: An Introduction to the Philosophy of Cognitive Science"**

Andy Clark

2000

Źródło

[4] **"Language, embodiment, and the cognitive niche"**

Andy Clark

2006 | Trends in Cognitive Sciences | DOI: 10.1016/j.tics.2006.06.012

Źródło

[5] **"Strong and Weak Emergence"**

David J. Chalmers

2008 | Oxford University Press eBooks | DOI: 10.1093/acprof:oso/9780199544318.003.0011

Źródło

[6] "Cognitive Ecology"

Edwin L. Hutchins

2010 | Topics in Cognitive Science | DOI: 10.1111/j.1756-8765.2010.01089.x

Źródło

[7] "Learning to navigate"

Edwin L. Hutchins

1993 | Cambridge University Press eBooks | DOI: 10.1017/cbo9780511625510.003

Źródło

[8] "The Distributed Cognition Perspective on Human Interaction"

Edwin Hutchins

2020 | Roots of Human Sociality | DOI: 10.4324/9781003135517-19

Źródło

[9] "Transactive Memory: A Contemporary Analysis of the Group Mind"

Daniel M. Wegner

1987 | Springer eBooks | DOI: 10.1007/978-1-4612-4634-3_9

Źródło

[10] "Ironic processes of mental control."

Daniel M. Wegner

1994 | Psychological Review | DOI: 10.1037//0033-295x.101.1.34

Źródło

[11] "Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips"

Betsy Sparrow, Jenny Liu, Daniel M. Wegner

2011 | Science | DOI: 10.1126/science.1207745

Źródło

[12] "A Computer Network Model of Human Transactive Memory"

Daniel M. Wegner

1995 | Social Cognition | DOI: 10.1521/soco.1995.13.3.319

Źródło

[13] "An empirical investigation of the impact of ChatGPT on creativity"

Byung Cheol Lee, Jaeyeon Chung

2024 | Nature Human Behaviour | DOI: 10.1038/s41562-024-01953-1

Źródło

[14] "Nonexpert Evaluations on Architectural Design Creativity Across Cultures"

Seung Wan Hong, Jae Seung Lee

2015 | Creativity Research Journal | DOI: 10.1080/10400419.2015.1087245

Źródło

[15] "Exposure-Structure Blending Network for High Dynamic Range Imaging of Dynamic Scenes"

Sang-Hoon Lee, Haesoo Chung, Nam Ik Cho

2020 | IEEE Access | DOI: 10.1109/access.2020.3005022

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: c4c3cc06-30dc-4c44-9f40-f53c116e8ac5