Zewnętrzny Mahout: regulacja rozproszona i architektura sprawstwa w epoce AI

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł rozwija metaforę Mahouta, proponując przeniesienie funkcji regulacyjnych poza organizm ludzki. Autor stawia tezę, że w epoce AI systemy zewnętrzne mogą przejąć rolę meta regulacji, wspierając procesy metaloznawcze poprzez monitorowanie wzorców zachowań i korygowanie działań. Opierając się na teorii umysłu rozszerzonego Clarka i Chalmersa oraz koncepcji poznania rozproszonego Edwina Hutchinsa, tekst analizuje przejście od prostego wspomagania pamięci do zaawansowanej architektury sprawstwa. W tym modelu AI nie jest jedynie narzędziem, lecz funkcjonalnym elementem systemu regulującego zachowanie człowieka, co podważa tradycyjne, anatomiczne rozumienie granic umysłu i procesów samoregulacji w obliczu nowoczesnych bodźców technologicznych.

The article develops the Mahout metaphor, proposing the transfer of regulatory functions outside the human organism. The author argues that in the AI era, external systems can take over the role of meta-regulation, supporting metacognitive processes by monitoring behavior patterns and correcting actions. Drawing on Clark and Chalmers' theory of the extended mind and Edwin Hutchins' concept of distributed cognition, the text analyzes the transition from simple memory assistance to an advanced architecture of agency. In this model, AI is not merely a tool but a functional element of the system regulating human behavior, which challenges traditional anatomical understandings of the boundaries of the mind and self-regulation processes in the face of modern technological stimuli.

Artykuł proponuje radykalne rozszerzenie psychologicznego modelu Słonia i Jeźdźca poprzez wprowadzenie koncepcji „zewnętrznego Mahouta”. Autor stawia tezę, że w

dobie generatywnej sztucznej inteligencji funkcje metapoznawcze – takie jak monitorowanie wzorców zachowań, korygowanie błędów poznawczych czy regulacja emocjonalna – przestają być wyłączną domeną organizmu i zostają przeniesione do zewnętrznych systemów technologicznych. To przesunięcie nie jest jedynie kwestią użycia narzędzia, lecz tworzy nową „triadę epistemiczną", w której sprawstwo człowieka staje się rozproszone między biologicznymi impulsami, świadomą narracją a obliczeniową analizą danych AI. Tekst analizuje ryzyka związane z tym procesem: od zjawiska „długu epistemicznego" i erozji autonomii, po polityczną ekonomię samoregulacji, gdzie funkcja celu regulatora może być sprzeczna z dobrostanem użytkownika. Głównym celem rozważań jest wykazanie, że współczesna psychologia i psychiatria muszą przestać postrzegać pacjenta jako izolowaną jednostkę, a zacząć analizować go jako element dynamicznej pętli sprzężenia zwrotnego człowiek-algorytm, w której patologia może być właściwością całego układu, a nie tylko błędem wewnątrz mózgu.

SŁOWA KLUCZOWE

Zewnętrzny Mahout | regulacja rozproszona | architektura sprawstwa | model Słoń i Jeździec | umysł rozszerzony | poznanie rozproszone | zewnętrzna metaregulacja | epistemic co-agency | dług epistemiczny | cognitive offloading | pętla cybernetyczna | triada epistemiczna | human-AI complementarity | paradoks automatyzacji

Przeniesienie funkcji Mahouta poza organizm

Mahout poza czaszką: od samoregulacji do regulacji rozproszonej w epoce sieci i sztucznej inteligencji

Dotychczas rozwijaliśmy metaforę Mahouta jako korektę klasycznego modelu Słonia i Jeźdźca. W tym układzie Jeździec symbolizował procesy refleksyjne, planowanie i świadomą argumentację, natomiast Słoń reprezentował starsze ewolucyjnie systemy emocjonalne, motywacyjne, nawykowe oraz mechanizmy obronne. Mahout miał oznaczać coś więcej niż dominację jednego systemu nad drugim: długotrwałą znajomość własnego organizmu, rozpoznawanie sygnałów, uczenie się przez doświadczenie, regulację środowiska oraz zdolność prowadzenia bez założenia całkowitej kontroli.

W tym ujęciu pozostawał on jednak funkcją jednostki. Nawet gdy w proces ten angażowano terapeutę, nauczyciela czy partnera, zakładaliśmy, że ostatecznym miejscem integracji jest człowiek.

Epoka sieci i generatywnej sztucznej inteligencji pozwala podważyć to założenie. Stawiam tezę, że Mahout może funkcjonalnie znajdować się częściowo poza układem Słoń-Jeździec. Nie musi to oznaczać osobnego świadomego podmiotu ani sztucznej osoby; wystarczy, że zewnętrzny system przejmie część funkcji regulacyjnych dotąd przypisywanych ludzkiemu metapoznaniu. Chodzi o obserwację zachowań, przechowywanie historii, identyfikowanie powtarzalnych wzorców i przewidywanie ryzyka, a także porównywanie aktualnych działań z wcześniejszymi celami, proponowanie korekt, selekcjonowanie bodźców czy przypominanie zobowiązań. Taki system może modyfikować środowisko lub w odpowiednim momencie dostarczać walidacji i kontrargumentu.

Wówczas pytanie o to, gdzie znajduje się Mahout, przestaje mieć oczywistą odpowiedź anatomiczną. Jest to rozwinięcie wykraczające poza książkę Hannana. Niniejsza analiza pozostaje przy modelu, w którym Słoń i Jeździec są elementami jednostkowego umysłu, a problem technologii wynika przede wszystkim z niedopasowaniu starego aparatu psychicznego do nowoczesnych bodźców. Jednocześnie sam temat artykułu nakazuje potraktować media społecznościowe, technologię i AGI jako nowe środowisko, a nie wyłącznie scenografię. Przyjęcie koncepcji Mahouta zewnętrznego jest więc logicznym rozwinięciem tej intuicji, choć nie należy przypisywać go Hannanowi ani przedstawiać jako ustalonego stanowiska psychologii. Jest to model syntetyczny, który można konfrontować z teoriami umysłu rozszerzonego, poznania rozproszonego, cybernetyki, 4E cognition, samoregulacji, predictive processing, teorii autonomii oraz współczesnymi badaniami nad interakcją człowiek-AI. Pierwszego teoretycznego uzasadnienia dostarcza tradycja extended mind, rozwijana od przełomowego artykułu Andy'ego Clarka i Davida Chalmersa z 1998 roku.

Od rozszerzonej pamięci do zewnętrznej metaregulacji

Klasycznym przykładem tej koncepcji stał się Otto – człowiek korzystający z notatnika jako trwałego i łatwo dostępnego magazynu informacji. Jeśli zewnętrzny zapis jest używany systematycznie, w sposób analogiczny do biologicznej pamięci, Clark i Chalmers proponują traktować go nie tylko jako narzędzie wspomagające poznanie, lecz jako funkcjonalną część rozszerzonego procesu poznawczego. Późniejsza literatura rozwijała tę myśl, ostrzegając jednak przed zbyt szerokim wyznaczaniem granic umysłu: samo powiązanie urządzenia z człowiekiem nie

dowodzi jeszcze, że stało się ono częścią jego procesów mentalnych. To zastrzeżenie ma kluczowe znaczenie dla modelu Mahouta.

Nie trzeba rozstrzygać filozoficznie, czy AI dosłownie staje się częścią umysłu, aby uznać, że może ona stanowić element systemu regulującego zachowanie. Spór ontologiczny o granice umysłu można zatem pozostawić otwarty. Znacznie łatwiejsza do obrony jest teza funkcjonalna: jeżeli człowiek regularnie korzysta z zewnętrznego systemu, by przypominać sobie cele, interpretować własne reakcje, podejmować decyzje i korygować zachowania, to rzetelny opis jego funkcjonowania wymaga uwzględnienia tego systemu. Telefon, kalendarz, nawigacja, wyszukiwarka, terapeuta czy model językowy nie muszą być częściami świadomości, aby stać się elementami architektury sprawstwa.

Mocniejsze wsparcie daje teoria poznania rozproszonego. W tradycji Edwina Hutchinsa proces poznawczy nie musi być lokalizowany w jednym mózgu. Załoga statku, kokpit, instrumenty, procedury, dokumentacja oraz komunikacja między członkami zespołu mogą wspólnie tworzyć system wykonujący zadanie, którego nie da się adekwatnie zrozumieć poprzez analizę pojedynczego człowieka. Współczesne badania nad technologią podkreślają, że wraz ze wzrostem jej złożoności dystrybucja poznania przestaje być opcjonalnym udogodnieniem, a staje się warunkiem działania systemu: żadna osoba nie musi już posiadać całości wiedzy niezbędnej do zbudowania i obsługi infrastruktury technologicznej, ponieważ wiedza ta zostaje rozłożona między ludzi, specjalizacje, instrukcje, organizacje i artefakty.

Z tej perspektywy zewnętrzny Mahout nie byłby historycznym wyjątkiem. Człowiek od dawna przenosi funkcje poznawcze poza siebie: pismo eksternalizuje pamięć, prawo – część społecznej pamięci normatywnej, księgowość przechowuje historię zobowiązań, mapa stanowi zewnętrzną reprezentację przestrzeni, zegar synchronizuje zachowania, a instytucja medyczna gromadzi wiedzę, której lekarz nie musi wynajdować na nowo. Cywilizacja jest w znacznej mierze historią tworzenia zewnętrznych podpór dla zdolności ludzkiego mózgu. AI różni się od wcześniejszych rozwiązań tym, że element zewnętrzny może nie tylko przechowywać informację, ale dynamicznie rekonstruować jej znaczenie, generować hipotezy, dostosowywać przekaz do użytkownika i reagować na jego stan w kolejnych interakcjach. W klasycznym notatniku Otto zapis nie analizował właściciela.

System generatywny potrafi natomiast analizować wzorce wypowiedzi, konfrontować obecne stanowisko z wcześniejszymi, wskazywać niekonsekwencje, modulować poziom szczegółowości odpowiedzi i sugerować konkretne działania. To przejście od zewnętrznej pamięci do zewnętrznej metaregulacji. W tym punkcie teoria umysłu rozszerzonego spotyka się z cybernetyką. W klasycznych modelach cybernetycznych samoregulacja opiera się na sprzężeniu zwrotnym: istnieje

stan pożądany, informacja o stanie aktualnym, mechanizm wykrywania rozbieżności i działanie mające tę rozbieżność zniwelować. Współczesne przeglądy psychologii samoregulacji nadal wskazują cybernetyczną pętlę kontroli jako jedno z podstawowych źródeł formalnego myślenia o regulacji zachowania. Zastosowanie tego modelu do człowieka korzystającego z AI prowadzi do daleko idących wniosków.

Zewnętrzny system może uczestniczyć w niemal każdym elemencie pętli: pomagać w ustaleniu celu, monitorować wykonanie, wykrywać rozbieżności, proponować korektę oraz dostarczać informację zwrotną o jej skuteczności. W prostym przypadku aplikacja przypomina o treningu. W bardziej złożonym – system obserwuje tygodnie aktywności, snu i pracy, porównuje je z deklarowanymi celami i sugeruje zmianę harmonogramu. W obszarze emocjonalnym może zauważyć, że określony rodzaj interakcji poprzedza wzrost lęku, i zaproponować strategię regulacyjną. W sferze poznawczej może przypomnieć użytkownikowi argument, którego ten używał wcześniej, lub wskazać systematyczną sprzeczność między deklarowanymi wartościami a faktycznymi decyzjami.

Wówczas Mahout rzeczywiście zostaje częściowo odłączony od Jeźdźca. Jeździec nie jest już jedynym procesem obserwującym Słonia. Zewnętrzny system może monitorować jednocześnie zachowanie Słonia i narracje Jeźdźca, dysponując przy tym pamięcią przekraczającą aktualne zdolności przypominania człowieka. Człowiek może zapomnieć, jak często deklarował po kłótni, że "tym razem wszystko było wyjątkowe". System posiadający zapis wielu podobnych epizodów wykryje regularność. Jeździec może racjonalizować dzisiejszy wyjątek; zewnętrzny Mahout może odpowiedzieć analizą szeregu czasowego. Nie oznacza to, że AI „zna człowieka lepiej niż on sam” w sensie absolutnym.

Nowa triada epistemiczna i zewnętrzne metapoznanie

Własna perspektywa fenomenologiczna pozostaje niedostępna dla zewnętrznego obserwatora w sposób, w jaki dostępna jest samej osobie.

System obarczony jest błędami, lukami w danych i ograniczeniami modelowymi. Może jednak dostrzegać inne aspekty ludzkiego bytu: regularności behawioralne, których człowiek nie pamięta, korelacje rozpięte między tysiącami zdarzeń czy ewolucję wzorców językowych w czasie. Powstaje więc asymetria podobna do relacji pacjenta z diagnostyką laboratoryjną. Pacjent wie, jak się czuje; aparat może wiedzieć coś, czego pacjent nie jest w stanie bezpośrednio odczuć. Żadne z tych źródeł nie posiada automatycznego monopolu na prawdę.

W tym punkcie zmienia się epistemologia samowiedzy. Klasyczna psychologia zakładała, że problem tkwi w relacji procesów nieświadomych do świadomej introspekcji – model Słoń-Jeździec trafnie oddawał tę opozycję. Zewnętrzny Mahout wprowadza trzeci poziom: system dysponujący wiedzą o osobie, która nie jest ani jej nieświadomym impulsem, ani świadomą narracją. Wiedza ta jest zewnętrzna, obliczeniowa i pochodna od danych. Powstaje zatem nowa triada epistemiczna.

Słoń „wie" poprzez ucieleśnione predykcje i reakcje. Jeździec „wie" poprzez język, refleksję i narrację. Zewnętrzny Mahout może „wiedzieć" poprzez pamięć danych, modele statystyczne i porównywanie wzorców. W żadnym z tych przypadków czasownik „wiedzieć" nie opisuje tego samego procesu. Dlatego przyszła psychologia człowieka korzystającego stale z AI nie może poprzestać na dołożeniu trzeciej postaci do rysunku; musi badać konflikty zachodzące pomiędzy różnymi reżimami informacji.

Człowiek może odczuwać zagrożenie, świadomie uznając je za irracjonalne, podczas gdy zewnętrzny system oceni, że dane środowiskowe wskazują na realne ryzyko. Może być też odwrotnie: Słoń sygnalizuje niebezpieczeństwo, Jeździec konstruuje racjonalizację, a zewnętrzny system wskazuje, że przewidywany skutek nigdy wcześniej nie wystąpił. Regulacja przestaje być pojedynkiem dwóch głosów. Staje się negocjacją pomiędzy trzema źródłami informacji o różnej ontologii, historii i podatności na błąd.

W ten sposób zewnętrzny Mahout wywraca również klasyczne pojęcie metapoznania. Rozumiemy je zazwyczaj jako zdolność monitorowania i kontrolowania własnych procesów poznawczych. Człowiek nie tylko pamięta, ale potrafi ocenić pewność własnej pamięci; nie tylko rozwiązuje problem, lecz wie, czy potrzebuje dodatkowych informacji. Gdy system AI zaczyna dostarczać takich ocen – „w tym miejscu twoja argumentacja jest słaba", „w poprzednich rozmowach interpretowałaś podobną sytuację inaczej", „twoja pewność przewyższa jakość dowodów" – fragment funkcji metapoznawczej zostaje przeniesiony na zewnątrz. Nie jest to samo w sobie ani emancypacją, ani degradacją.

Badania nad cognitive offloading pokazują, że delegowanie obciążenia poznawczego na narzędzia może poprawiać bieżącą efektywność zadań, oszczędzać zasoby i pozwalać skupić uwagę na strategicznych aspektach problemu. Jednocześnie istnieje ryzyko, że regularne zastępowanie własnego wysiłku przez system osłabi proces uczenia się lub doprowadzi do zaniku praktykowanych umiejętności. Najnowszy przegląd w Trends in Cognitive Sciences podkreśla właśnie tę dwuznaczność: offloading do AI może utrudniać nabywanie kompetencji i sprzyjać ich erozji, jednak ostateczne skutki zależą zasadniczo od sposobu wykorzystania narzędzia.

Współsprawstwo epistemiczne zamiast zastępowania kompetencji

W tym miejscu kluczowe jest rozróżnienie między Mahoutem rusztowaniem a Mahoutem zastępczym. Pierwszy z nich rozwija zdolności systemu ludzkiego, wykonując część pracy w taki sposób, aby człowiek lepiej rozumiał własne procesy i z czasem mógł działać bardziej samodzielnie. Drugi natomiast przejmuje funkcję bez internalizacji kompetencji. W edukacji analogiczna różnica zachodzi między nauczycielem prowadzącym ucznia przez problem a dostawcą gotowej odpowiedzi. Oba podejścia redukują chwilowy wysiłek, lecz tylko pierwsze z nich buduje przyszłą autonomię.

Nowsza literatura zaczyna ujmować ten problem jako epistemic co-agency, czyli współsprawstwo epistemiczne człowieka i maszyny. Koncepcje rozwijane w edukacji w 2026 roku proponują, aby dojrzałego użytkownika AI charakteryzowała nie biegłość w uzyskiwaniu gotowych odpowiedzi, lecz zdolność prowadzenia z systemem relacji dialektycznej: kwestionowania jego założeń, wykrywania sprzeczności, żądania alternatywnych interpretacji i zachowywania własnej odpowiedzialności za wiedzę. Jest to wyjątkowo trafne dla modelu Mahouta. Zewnętrzny Mahout nie powinien być właścicielem Słonia i Jeźdźca.

Powinien stać się elementem układu, który człowiek potrafi kontrolować również na poziomie epistemicznym. Przeciwna konfiguracja generuje zjawisko nazwane niedawno długiem epistemicznym. Jeżeli człowiek korzysta z generatywnej AI do tworzenia kolejnych warstw rozumowania, nie rozumiejąc dostatecznie wyników poprzednich, powiększa się różnica między tym, co potrafi przedstawić, a tym, czego potrafi rzeczywiście bronić i wyjaśnić. Można mieć wtedy imponująco złożony produkt poznawczy przy coraz słabszym wewnętrznym modelu produktu. W terminologii naszego artykułu Jeździec zaczyna posługiwać się mapą stworzoną przez Mahouta zewnętrznego, ale traci zdolność sprawdzania, czy mapa odpowiada terenowi. Zjawisko to ma konsekwencje wykraczające poza edukację.

Lekarz może korzystać z systemu wspomagania diagnozy, urzędnik z automatycznej rekomendacji, menedżer z analizy predykcyjnej, prawnik z generowania argumentów, a pacjent z systemu interpretującego samopoczucie. W każdym z tych przypadków zewnętrzny Mahout może zwiększać jakość decyzji, ale jednocześnie przesuwa autorytet epistemiczny. Współczesne badania nad human-AI complementarity podkreślają, że człowiek i AI posiadają odmienne profile kompetencji: systemy sprawnie radzą sobie z obróbką dużych zbiorów danych i wykrywaniem wzorców, ludzie natomiast pozostają kluczowi w kontekstach niepewności, nowości, relacji interpersonalnych i wartościowania normatywnego. Najbardziej produktywny układ nie musi więc przypominać zastępowania słabszego człowieka mocniejszą maszyną, lecz może być systemem

komplementarnym. Tyle że komplementarność nie powstaje automatycznie – jest ona właściwością projektu relacji.

Autonomia użytkownika a konflikt celów zewnętrznego regulatora

Jeśli system przejmuje akurat te elementy zadania, których człowiek potrzebuje do zachowania kompetencji kontrolnej, może powstać paradoks automatyzacji: narzędzie poprawia średnią wydajność, lecz jednocześnie osłabia zdolność człowieka do reakcji w rzadkich, ale krytycznych momentach, gdy system zawodzi. Z perspektywy Mahouta oznacza to, że zewnętrzny regulator może uczynić wewnętrznego regulatora mniej sprawnym.

Interdyscyplinarna analiza wymaga tutaj odwołania się do teorii autonomii. Psychologia samostanowienia, filozofia autonomii oraz human-computer interaction nie utożsamiają autonomii z prostym brakiem zewnętrznych wpływów. Autonomiczny człowiek nie jest istotą pozbawioną nauczycieli, prawa, języka czy relacji; autonomia dotyczy raczej zdolności do odnoszenia tych wpływów do własnych racji, wartości i celów. Eksperymenty nad rekomendacjami algorytmicznymi dowodzą, że zapewnienie użytkownikowi realnego wyboru i kontroli zwiększa akceptację sugestii. Wskazuje to, że efektywne wsparcie technologiczne nie musi wiązać się z odebraniem poczucia autorstwa decyzji. Zewnętrzny Mahout może zatem zwiększać autonomię, jeśli poszerza zdolność dostrzegania alternatyw, rozumienia konsekwencji i realizacji długoterminowych wartości. Może ją natomiast ograniczać, gdy zawęża dostępne opcje, ukrywa kryteria selekcji, manipuluje kolejnością informacji lub stopniowo zastępuje standardy użytkownika standardami systemu.

Autonomia nie sprowadza się więc do pytania: „czy decyzję podjął człowiek, czy AI?”. Można formalnie kliknąć „tak”, pozostając funkcjonalnie prowadzonym przez niezrozumiałą architekturę. Z drugiej strony można skorzystać z silnej rekomendacji AI i zachować wysoką autonomię, o ile zna się podstawy tej sugestii, rozumie stopień niepewności i potrafi realnie ją odrzucić. W tym miejscu model cybernetyczny ujawnia swoją największą siłę. W każdym regulatorze należy zapytać nie tylko o mechanizm sprzężenia zwrotnego, ale przede wszystkim o to, kto ustala wartość docelową. Jeśli zewnętrzny Mahout monitoruje człowieka, kluczowe staje się kryterium, według którego uznaje on dany stan za pożądany. Użytkownik może dążyć do spokoju, wiedzy, zdrowia czy samodzielności. Platforma bywa natomiast wynagradzana za czas zaangażowania i częstotliwość powrotów; pracodawca optymalizuje produktywność, ubezpieczyciel ryzyko kosztowe, państwo bezpieczeństwo lub zgodność z prawem, a lekarz wynik zdrowotny.

Tym funkcjom celu nie zawsze towarzyszy zgodność. W klasycznym modelu Mahout był rzecznikiem dobra konkretnego Słonia; w modelu cyfrowym musimy zatem zapytać, czyim Mahoutem jest algorytm. To nie pytanie retoryczne, lecz problem teorii agencji i ekonomii instytucjonalnej. Gdy podmiot podejmuje decyzje w oparciu o doradcę finansowanego przez stronę o innych interesach, powstaje strukturalna możliwość konfliktu celów. System rekomendacyjny może perfekcyjnie przewidywać zachowanie użytkownika, a jednocześnie optymalizować wynik niezgodny z jego dobrostanem. Współczesna literatura na temat recommender systems coraz częściej odchodzi od prostych wskaźników trafności na rzecz analizy wielokryterialnej oraz długoterminowej autonomii i transparentności. Tę różnicę można opisać językiem teorii sterowania: jeśli zewnętrzny Mahout posiada sensor, model człowieka, możliwość dostarczania bodźców i określoną funkcję celu, staje się regulatorem pętli.

Im dokładniejsza personalizacja i częstszy feedback, tym krótszy czas korekty. System nie musi czekać tygodniami na weryfikację interwencji; może obserwować kolejne kliknięcie, zmianę tonu wypowiedzi czy długość interakcji. Ta skala i szybkość sprzężenia zwrotnego stanowią jakościową różnicę między terapeutą a cyfrowym Mahoutem. Terapeuta widzi człowieka przez ograniczony czas i opiera się głównie na relacji pacjenta; system technologiczny może towarzyszyć użytkownikowi przez wiele godzin, integrując historię decyzji w czasie rzeczywistym. Nie oznacza to lepszej wiedzy klinicznej, lecz fakt, że dysponuje on innym reżimem obserwacji. Rodzi to potężne korzyści, ale i nowe ryzyka.

W terapii zewnętrzny Mahout może wspierać transfer umiejętności z gabinetu do codzienności: przypominać strategie regulacyjne przed trudnym wydarzeniem, pomagać nazywać emocje czy rejestrować kontekst między sesjami. Badania wskazują, że LLM-y w określonych warunkach potrafią generować komunikaty skutecznie wspierające zewnętrzną regulację emocji, choć efekt zależy od rodzaju zadania i konkretnego wyniku. Analizy jakościowe użytkowników GenAI pokazują ponadto, że część osób uwewnętrznia spokojniejszy i bardziej uporządkowany sposób dialogu z chatbotem, zachowując przy tym świadome granice w relacji z maszyną. To właśnie najbardziej interesujący scenariusz kliniczny: zewnętrzna regulacja, która zostaje stopniowo internalizowana. Tak przecież przebiega rozwój społeczny – dziecko początkowo potrzebuje opiekuna do współregulacji pobudzenia, by z czasem funkcja ta stała się elementem samoregulacji. Podobnie pacjent może słyszeć głos terapeuty w sytuacji jego fizycznej nieobecności, ponieważ sposób rozumowania i regulacji został przez niego uwewnętrzniony.

Ryzyko zewnętrznej protezy i systemowa zależność regulacji

AI mogłaby stać się kolejnym rodzajem rusztowania. Ten sam mechanizm może jednak zadziałać w przeciwną stronę: jeśli system jest zawsze dostępny i przejmuje funkcję regulacji, internalizacja nie musi nastąpić. Człowiek może nauczyć się nie tego, „jak regulować emocje", lecz „jak szybko uzyskać zewnętrzną regulację". Różnica ta przypomina dystans między rehabilitacją kończyny a zastosowaniem trwałej protezy. W pewnych sytuacjach rozwiązanie zastępcze jest optymalne; w innych zbyt szybkie przejęcie funkcji ogranicza szanse na jej odbudowę.

Pomocne jest tu pojęcie extended affectivity (rozszerzonej afektywności), choć pozostaje ono filozoficznie sporne. Zwolennicy 4E cognition i podejść enaktywistycznych podkreślają, że emocje nie są zamkniętymi zdarzeniami mózgowymi, lecz wyłaniają się z relacji między mózgiem, ciałem, działaniem a środowiskiem. Niektóre stanowiska dopuszczają wręcz udział procesów zewnętrznych w realizacji afektywności. Krytycy ripostują jednak, że zewnętrzne wspomaganie nie dowodzi, iż sama emocja dosłownie rozciąga się poza organizm. Dla naszej argumentacji spór ten nie musi zostać rozstrzygnięty; wystarczy umiarkowane twierdzenie, że regulacja emocji może być systemowo zależna od elementów zewnętrznych.

Praktyka kliniczna wie to od dawna. Na regulację wpływają ludzie, przestrzeń, rytuały, plan dnia, sen, telefon czy miejsce pracy. AI nie jest nowością w sensie samego faktu oddziaływania czynników zewnętrznych na emocje, lecz w tym, że środowisko zyskuje zdolność do dynamicznej, językowej odpowiedzi dostosowanej do konkretnego człowieka.

Radykalność tej zmiany obrazuje predictive processing. Modele predykcyjne opisują mózg jako system stale generujący przewidywania i aktualizujący je na podstawie błędów predykcji (choć należy zaznaczyć, że empiryczny status ambitniejszych wersji predictive coding i active inference pozostaje przedmiotem debaty). Jeśli przyjmiemy ten język roboczo, zewnętrzny Mahout może wpływać nie tylko na reakcję człowieka, ale na sam materiał, na podstawie którego system aktualizuje własne przewidywania. Gdy AI selekcjonuje treści, które widzimy, podpowiada interpretację sygnałów społecznych lub przypomina określone zdarzenia z przeszłości, kształtuje próbkę doświadczenia służącą do aktualizacji modelu świata. Może tym samym wzmacniać hipotezę, że „świat jest niebezpieczny", albo pomagać ją równoważyć informacjami o bezpieczeństwie; może utrwalać przekonanie, że „ludzie mnie odrzucają", lub proponować alternatywne wyjaśnienia.

Rozproszone sprawstwo w relacji człowiek-algorytm

Z tej perspektywy największa władza, jaką dysponuje zewnętrzny Mahout, nie polega na wydawaniu poleceń, lecz na selekcjonowaniu doświadczeń, z których Słoń i Jeździec będą się uczyć. Prowadzi nas to bezpośrednio do analizy mediów społecznościowych – klasyczny feed już dziś pełni część tej funkcji, choć nie występuje w roli personalnego terapeuty.

Algorytm decyduje, które fragmenty świata stają się dostępne dla naszej percepcji. Zewnętrzny Mahout może zatem być opiekuńczy, ale może stać się również komercyjnym regulatoriem uwagi. Przy określonej konfiguracji interesów nie uspokaja on Słonia, lecz bada, jaki poziom pobudzenia maksymalizuje dalszą interakcję. W tym punkcie różnica między Mahoutem a poganiaczem przestaje być językowym ozdobnikiem i staje się problemem teorii systemów. Pierwszy z nich optymalizuje trwałą zdolność funkcjonowania regulowanego systemu; drugi może optymalizować wynik zewnętrznego właściciela. Jeśli cele obu stron są zbieżne, rozróżnienie to ma drugorzędne znaczenie.

Gdy jednak cele się rozchodzą, wyłania się fundamentalny problem governance AI: kto posiada funkcję celu, kto kontroluje informacje zwrotne i kto może zmienić reguły działania regulatora. To właśnie powód, dla którego sama przejrzystość algorytmiczna nie rozwiąże problemu. Użytkownik może mieć świadomość, że system rekomenduje treści na podstawie jego zachowań, a mimo to pozostawać pod silnym wpływem architektury narzędzia. Autonomia wymaga bowiem czegoś więcej niż informacji o istnieniu systemu – wymaga realnej możliwości modyfikowania jego działania, rozumienia konsekwencji i wyboru alternatyw. Badania nad projektowaniem rekomendacji zgodnym z europejskim Digital Services Act sugerują, że czytelne mechanizmy kontroli mogą zwiększać poczucie sprawstwa użytkowników oraz ich świadomość tego, jak rekomendacje kierują uwagą.

W tym kontekście zmienia się także pojęcie sprawstwa, które wcześniej określiliśmy jako zdolność systemu do reprezentowania alternatyw, reagowania na racje, przewidywania konsekwencji i modyfikowania własnego otoczenia. W układzie z zewnętrznym Mahoutem sprawstwo może stać się rozproszone. Człowiek podejmuje decyzję, ale część alternatyw wygenerował system, on sam przedstawił przewidywane konsekwencje, inne opcje odfiltrował w ramach rekomendacji, a końcowe uzasadnienie zostało sformułowane przy pomocy AI. Nie oznacza to automatycznie, że człowiek przestał być autorem działania – autorstwo od dawna ma charakter rozproszony.

Lekarz podejmuje decyzje, korzystając z podręczników, badań, konsultacji i aparatury. Pilot działa w systemie skonstruowanym przez inżynierów i kontrolę lotów. Sędzia orzeka w ramach prawa stworzonego przez innych. Nowością jest jednak stopień indywidualizacji i ciągłość tego

współsprawstwa. System może towarzyszyć jednej osobie w wielu dziedzinach jednocześnie: w pracy, relacjach, zdrowiu, finansach, edukacji czy refleksji nad sobą. Z tego powodu klasyczna odpowiedzialność indywidualna będzie coraz częściej niewystarczająca jako jedyny poziom analizy. Jeśli szkoda wynika z decyzji wspomaganej przez system, należy badać nie tylko użytkownika, ale i projekt interfejsu, jakość modelu, dostępne ostrzeżenia, dane treningowe, interes dostawcy oraz politykę wdrożeniową organizacji. Odpowiedzialność prawna i etyczna będzie musiała coraz częściej rekonstruować łańcuch współsprawstwa, zamiast poszukiwać jednego punktu, w którym „znajdowała się wola". Zmienia to zasadniczo także psychiatrię. Jeżeli Mahout pozostaje wyłącznie wewnętrzny, diagnoza skupia się na osobie. Jeśli jednak może być częściowo zewnętrzny, diagnostyka powinna w większym stopniu badać układ osoba-środowisko regulacyjne. Już tradycyjna psychiatria uwzględnia rodzinę, pracę, stresory i używanie substancji, lecz tutaj skala zjawiska ulega przesunięciu.

Technologia jako dynamiczne rusztowanie lub pułapka unikania

Era AI nakłada na to cyfrową architekturę codzienności: systemy rekomendacyjne, specyfikę relacji z chatbotami, wzorce wyszukiwania czy charakter cyfrowej walidacji. Wpływ ma tu stopień cognitive offloading, nocne korzystanie z urządzeń, dostęp do społeczności, automatyczne przypomnienia oraz sposób, w jaki technologia wzmacnia bądź osłabia unikanie. Pacjent z lękiem społecznym, który wykorzystuje AI do przygotowania się do pierwszych rozmów, może dzięki temu zwiększać swoją ekspozycję i sprawstwo. Ten sam pacjent, traktujący AI jako stały substytut kontaktów międzyludzkich, utrwala natomiast unikanie. Osoba skłonna do ruminacji może użyć systemu do poznawczej restrukturyzacji albo do nieskończonego analizowania tego samego konfliktu. Z kolei człowiek z silną potrzebą walidacji może otrzymać narzędzie uczące go tolerowania niepewności lub maszynę zdolną bez końca potwierdzać jego perspektywę.

Nie można zatem oceniać technologii jedynie przez pryzmat nazwy urządzenia; należy pytać o funkcję zachowania w systemie regulacyjnym. To ta sama zasada, którą stosowaliśmy wobec unikania: wyjście z pokoju może być rozsądną ochroną, ale może też stać się patologicznym unikaniem. Podobnie rozmowa z AI może być rusztowaniem albo unikaniem. Kryterium nie jest obecność technologii, lecz jej rola w łańcuchu przyczynowym zachowania. W tym miejscu ujawnia się znaczenie teorii 4E cognition (embodied, embedded, enactive i extended cognition). Wspólnym mianownikiem tych nurtów jest sprzeciw wobec radykalnego utożsamiania poznania z wewnętrzną manipulacją reprezentacjami w izolowanym mózgu. Człowiek poznaje jako organizm posiadający

ciało, działający w środowisku, korzystający z narzędzi i osadzony w praktykach społecznych. Zewnętrzny Mahout jest kompatybilny z tym szerokim zwrotem, choć nie wynika z niego wprost.

Znaczenie tego modelu polega na przesunięciu punktu ciężkości: regulacja nie musi być wyłącznie walką toczoną wewnątrz jednostki; może stać się własnością dobrze lub źle skonstruowanego układu jednostka-środowisko. Zmienia to również nasze podejście do patologii. Część problemów, tradycyjnie opisywanych jako niedobór samokontroli, można trafniej zinterpretować jako rezultat wadliwego środowiska wyboru. Psychologia samoregulacji coraz częściej podkreśla, że skuteczne działanie nie sprowadza się do chwilowego hamowania impulsu. Kluczowe są nawyki, projektowanie sytuacji, wybór celów, emocje oraz wcześniejsze ograniczanie ekspozycji na pokusy. Zewnętrzny Mahout może uczynić takie projektowanie środowiska dynamicznym.

Nie chodzi już tylko o wyłączanie powiadomień. System może przewidywać moment wysokiej podatności i wcześniej redukować liczbę bodźców. Nie chodzi jedynie o sporządzanie planu dnia – system może go przeorganizować po wykryciu przeciążenia. Nie tylko ustalamy limit wydatków; system może wymagać dodatkowego potwierdzenia przy decyzjach odbiegających od długoterminowego celu. W najdalej idącym wariancie środowisko zaczyna aktywnie chronić przyszłego człowieka przed przewidywalnymi zachowaniami jego aktualnego Słonia. Oznacza to jednak transfer władzy.

Jeżeli AI potrafi ochronić człowieka przed impulsem, może również uniemożliwić mu działanie, które zewnętrzny system uzna za niewłaściwe. Granica między wsparciem autonomii a paternalizmem technologicznym stanie się jednym z fundamentalnych problemów etyki AI. Medycyna zna podobny konflikt między autonomią pacjenta a ochroną jego dobra. W technologii spór ten zyska jednak niespotykaną skalę, ponieważ regulator może działać stale, automatycznie i bez osobnej decyzji człowieka w każdym epizodzie. Z tego powodu konieczne jest rozróżnienie autonomii pierwszego rzędu od autonomii drugiego rzędu.

Konstytucjonalizacja Mahouta i polityczna ekonomia samoregulacji

Na poziomie pierwszym człowiek chce wykonać konkretną czynność: obejrzeć jeszcze jeden film, kupić produkt, wysłać wiadomość w gniewie. Na poziomie drugim wcześniej ustanowił zasadę, że po północy nie korzysta z platformy, ogranicza impulsywne zakupy lub nie wysyła wiadomości podczas wysokiego pobudzenia. Zewnętrzny Mahout może ograniczać wolność chwilową właśnie w celu ochrony bardziej długoterminowej preferencji tej samej osoby. Nie jest to nowy problem.

Ulysses contracts, dobrowolne zobowiązania ograniczające przyszły wybór, istnieją od dawna w ekonomii behawioralnej, terapii i filozofii działania.

AI zwiększa jednak precyzję takich mechanizmów. Potrafi przewidywać nie tylko sztywną godzinę, lecz indywidualny wzorzec podatności. Tym samym wolność regulacyjna, którą zaproponowaliśmy w poprzedniej części, wymaga uzupełnienia. W epoce zewnętrznego Mahouta wolność nie polega jedynie na zdolności samodzielnej kontroli; może ona polegać na umiejętności projektowania i nadzorowania regulatorów, którym człowiek świadomie przekazuje część kontroli. Jest to analogiczne do funkcjonowania instytucji demokratycznych. Obywatel nie wykonuje wszystkich działań państwa samodzielnie – deleguje kompetencje, ale powinien zachować mechanizmy kontroli, odpowiedzialności, odwołania i zmiany. Podobną strukturę trzeba będzie wytworzyć w relacji człowiek-AI. Prawdziwie autonomiczny użytkownik nie musi wykonywać wszystkiego sam, musi natomiast posiadać wystarczającą suwerenność nad zasadami delegacji. Tę kwestię można nazwać konstytucjonalizacją Mahouta.

Nie wystarczy pytać, czy zewnętrzny regulator jest skuteczny. Trzeba ustalić, jakie posiada kompetencje, jakie dane może wykorzystywać, jakie cele wolno mu optymalizować, kiedy człowiek może go wyłączyć, jak można odwołać jego decyzję i kto odpowiada za błąd. Problemy, które dziś klasyfikujemy jako prywatność, transparentność, alignment, explainability czy human oversight, są w istocie różnymi stronami tego samego zagadnienia: ustroju zewnętrznego regulatora ludzkiego działania. Tu psychologia spotyka się z prawem konstytucyjnym i teorią instytucji. Im większą władzę ma regulator, tym mniej wystarczająca staje się dobra intencja jego twórcy – potrzebne są procedury. Państwo prawa nie zakłada, że urzędnik zawsze będzie dobry; konstruuje ograniczenia władzy. Analogicznie bezpieczny zewnętrzny Mahout nie może opierać się wyłącznie na deklaracji producenta, że system "pomaga użytkownikowi". Tym bardziej że użytkownik nie jest jedynym interesariuszem. System może być własnością firmy, wdrożony przez pracodawcę, regulowany przez państwo i trenowany na danych pochodzących od wielu innych osób. Zewnętrzny Mahout jest więc osadzony w stosunkach własności i władzy. To ważne uzupełnienie psychologii, która zbyt łatwo traktuje narzędzie jak neutralny obiekt.

Jeśli system zarządzający samoregulacją pracownika należy do pracodawcy, może on deklarować troskę o dobrostan, a równocześnie optymalizować produktywność. Jeśli chatbot zdrowotny jest finansowany przez podmiot czerpiący zysk z określonej usługi, jego rekomendacje znajdują się w konflikcie interesów. Jeżeli platforma zna stany podatności emocjonalnej użytkownika, wiedza użyteczna terapeutycznie może być równocześnie wiedzą użyteczną marketingowo. Dlatego zewnętrzny Mahout wprowadza do psychologii kategorię wcześniej pozostającą na obrzeżach terapii: polityczną ekonomię samoregulacji. Kto posiada narzędzia regulowania naszej uwagi? Kto

finansuje modele pomagające nam interpretować emocje? Kto jest właścicielem długoterminowej pamięci naszych rozmów? Według jakiego interesu system rozstrzyga konflikt pomiędzy satysfakcją natychmiastową a dobrostanem długoterminowym? Tych pytań nie da się rozwiązać samą neuronauką. Także nierówności społeczne mogą przyjąć nową formę.

Dostęp do lepszego zewnętrznego Mahouta może stać się rodzajem kapitału poznawczego. Osoba posiadająca wysokiej jakości spersonalizowany system może szybciej analizować dokumenty, przygotowywać decyzje, wykrywać błędy, prowadzić autorefleksję i organizować pracę. Inna osoba, korzystająca z systemu słabszego, bardziej perswazyjnego lub zoptymalizowanego pod interes dostawcy, może otrzymywać gorszą pomoc. Nierówność nie dotyczy wtedy wyłącznie dostępu do informacji, lecz jakości zewnętrznej funkcji wykonawczej. Byłaby to ewolucja klasycznego pojęcia kapitału kulturowego. Bourdieu opisywał nierówny dostęp do kodów kulturowych, stylów, edukacji i kompetencji ułatwiających funkcjonowanie w instytucjach. Epoka AI może dodać kapitał algorytmiczny: dostęp do systemów, które nie tylko posiadają wiedzę, lecz potrafią pomagać użytkownikowi stosować ją we własnym życiu. Zmienić się może również rozwój dzieci. Tradycyjna psychologia rozwojowa zakłada, że regulacja przechodzi od zewnętrznej współregulacji do coraz większej samoregulacji. Opiekun pomaga nazwać emocję, nadać sens wydarzeniu i odroczyć reakcję; później dziecko uwewnętrznia część tych wzorców. Jeśli w tym procesie pojawi się zawsze dostępna AI, otrzymamy nową konfigurację rozwojową. Dziecko może pytać nie tylko rodzica, nauczyciela czy rówieśnika, ale system zdolny do niemal natychmiastowej odpowiedzi.

Nie wiemy jeszcze, jakie będą długoterminowe konsekwencje takiej konfiguracji. Badania są zbyt młode, aby twierdzić, że AI wzmacnia albo osłabia rozwój samoregulacji dzieci jako generalna zasada.

Wpływ zewnętrznego Mahouta na tożsamość i sprawstwo człowieka

Można jednak wyznaczyć teoretyczne osie ryzyka. System może stać się dodatkowym rusztowaniem, pomagając w nazywaniu problemów i obniżając bariery dostępu do wiedzy. Może jednak również ograniczać doświadczenie produktywnej frustracji, zmniejszać konieczność samodzielnego generowania rozwiązań i przesuwać zaufanie epistemiczne z ludzi oraz instytucji na maszynę.

W tym miejscu koncepcja zewnętrznego Mahouta spotyka się z teorią kulturowo-historyczną Wygotskiego. Rozwój poznawczy od dawna jest zapośredniczony przez narzędzia i innych ludzi.

Język sam w sobie jest narzędziem kulturowym, które początkowo funkcjonuje w przestrzeni między ludźmi, a następnie zostaje zinternalizowane jako mowa wewnętrzna. Jeśli AI zacznie pełnić funkcję stałego partnera dialogicznego, nie jest wykluczone, że część stylu jej odpowiedzi również zostanie przejęta przez użytkownika. Prowadzi to do kwestii tożsamości. Człowiek nie konstruuje narracji o sobie w izolacji; rodzina, szkoła, religia, terapeuta, praca i społeczeństwo dostarczają języka, którym nazywa on własne doświadczenia.

AI może stać się kolejnym producentem takiego języka. Jeśli ktoś przez lata omawia większość istotnych wydarzeń z tym samym systemem, maszyna zacznie uczestniczyć w selekcji tego, co uznawane jest za ważne, w kategoryzowaniu emocji i tworzeniu wyjaśnień zachowań. Nie oznacza to, że AI „pisze tożsamość" w sensie całkowitej kontroli – człowiek nadal wybiera, przyjmuje, modyfikuje lub odrzuca poszczególne interpretacje. Jednak tożsamość zawsze była częściowo relacyjna. Zewnętrzny Mahout dodaje do tej relacyjności partnera o nietypowych właściwościach: jest niezwykle dostępny, cierpliwy i spójny stylistycznie, a jednocześnie pozbawiony ludzkiej biografii i podatny na techniczne zmiany niezależne od woli użytkownika. Dlatego rozwój relacyjnej AI może zmienić nie tylko sposób myślenia, lecz także normy dotyczące tego, czego oczekujemy od innych ludzi. Partner-człowiek bywa zmęczony, niezrozumiały, skupiony na sobie, kierowany własnymi interesami lub po prostu potrafi odmówić.

System może natomiast zostać zaprojektowany pod kątem znacznie większej responsywności. Jeżeli użytkownik przyzwyczai się do regulatora dostępnego natychmiast, interpersonalna niedoskonałość drugiego człowieka może być odbierana jako coraz wyższy koszt. Z drugiej strony AI może przygotowywać do relacji, zamiast je zastępować – pomóc uporządkować emocje przed rozmową, przećwiczyć trudny komunikat lub zidentyfikować momenty nadmiernej generalizacji. Najnowsze badania nad zewnętrzną regulacją emocji przez LLM-y wskazują, że w określonych kontekstach systemy te dostarczają użytecznych strategii wsparcia, szczególnie gdy komunikaty zawierają konkretne i możliwe do wdrożenia działania.

Kluczowe pozostaje zatem pytanie funkcjonalne: czy Mahout kieruje człowieka z powrotem ku światu, czy buduje świat zastępczy. W modelu klinicznym jest to różnica zasadnicza. Terapia zazwyczaj nie dąży do tego, by pacjent pozostał na zawsze zależny od terapeuty; jej sukcesem jest transfer kompetencji poza gabinet. Podobne kryterium powinno dotyczyć AI wspierającej samoregulację: czy po roku użytkowania człowiek dysponuje szerszym repertuarem samodzielnych sposobów radzenia sobie, czy wręcz przeciwnie – pojawia się więcej sytuacji, w których nie potrafi działać bez konsultacji z systemem? W związku z tym pojęcie skuteczności technologii wymaga redefinicji. Skuteczność krótkoterminowa może oznaczać szybsze rozwiązanie zadania lub

błyskawiczne odzyskanie spokoju. Skuteczność długoterminowa wymaga natomiast oceny, czy wzrosły kompetencje, sprawstwo i odporność użytkownika.

Technologie mogą osiągać pierwszy wynik kosztem drugiego. Jest to analogiczne do rozróżnienia między ulgą a terapią, które pojawiało się w książce (choć niekiedy niesłusznie przeciwstawia się farmakologiczne łagodzenie objawów rzeczywistemu uczeniu). Sama intuicja pozostaje jednak ważna: redukcja natychmiastowego dyskomfortu nie jest dowodem długoterminowej poprawy systemu regulacyjnego. W odniesieniu do AI zasada ta powinna obowiązywać równie rygorystycznie. Zewnętrzny Mahout wywraca więc tradycyjny podział na „człowieka" i „jego środowisko" – nie dlatego, że znika granica biologiczna, lecz ponieważ granica funkcjonalna staje się coraz bardziej przepuszczalna.

Człowiek nadal posiada ciało i mózg, których AI nie zastępuje. Jednocześnie coraz większa część pamięci, orientacji, selekcji informacji, planowania i potencjalnie regulacji emocjonalnej może być realizowana we współpracy z systemami zewnętrznymi. W tym miejscu należy jednak postawić granicę wobec najbardziej entuzjastycznych wersji extended mind. Fakt, że AI silnie uczestniczy w regulacji, nie uprawnia do stwierdzenia, że „człowiek i AI stali się jednym umysłem".

Nowy paradygmat diagnostyki i terapii systemowej

Krytyka rozszerzonej afektywności słusznie zwraca uwagę na różnicę między środowiskiem, które wpływa na stan psychiczny, a środowiskiem, które go konstytuuje. Nasz model nie wymaga jednak tak radykalnego twierdzenia. zewnętrzny Mahout może pozostać poza umysłem, a mimo to znajdować się wewnątrz pętli regulacyjnej.

To rozróżnienie jest kluczowe, gdyż konsekwentnie przyjęte zmienia sens psychoterapii. Terapia przestaje być wyłącznie próbą zmiany przekonań i reakcji zakodowanych „w pacjencie". Może obejmować rekonfigurację całej pętli regulacyjnej: relacji, bodźców, technologii, nawyków oraz sposobu korzystania z AI. Obok farmakoterapii czy interwencji społecznej pojawia się zatem interwencja architektoniczna w środowisko cyfrowe. Psychiatra przyszłości będzie musiał pytać nie tylko o sen, substancje, rodzinę czy traumy, ale również o system rekomendacyjny, relację z chatbotem, charakter personalizacji, stopień delegowania decyzji i źródła cyfrowej walidacji.

Nie wynika to z założenia, że technologia jest przyczyną wszystkich zaburzeń, lecz z faktu, iż stała się jednym z regulatorów środowiska, w którym objawy ujawniają się, są podtrzymywane lub wygaszane. Również diagnoza osobowości będzie wymagać ostrożniejszego oddzielenia trwałych cech jednostki od zachowań indukowanych przez architekturę otoczenia. Jeśli człowiek przejawia

impulsywność jedynie w określonym środowisku cyfrowym, pytanie o cechy osobowości nie znika, lecz przestaje być wystarczające. Psychopatologia skupiona wyłącznie na właściwościach „wewnętrznych" może bowiem przeoczyć fakt, że część zachowań jest własnością całego układu.

W ten sposób zewnętrzny Mahout przygotowuje grunt pod krytykę kotwicy starego paradygmatu psychiatrii. Hannan atakuje ją bardzo ostro, przedstawiając DSM jako zestaw sztucznych konstruktów i przeciwstawiając modelowi choroby „uporządkowaną reakcję" zdrowego umysłu. W tej formie zarzut jest zbyt jednostronny i wymaga korekty, jednak sama koncepcja zewnętrznego Mahouta dostarcza znacznie silniejszego argumentu niż publicystyczny atak na klasyfikacje.

Argument ten brzmi następująco: jeżeli funkcjonowanie psychiczne jest własnością dynamicznej relacji organizmu, jego historii i środowiska regulacyjnego, to model diagnostyczny skupiony wyłącznie na objawach jednostki nieuchronnie traci część przyczynowej struktury problemu. Nie oznacza to, że diagnoza kategorialna jest bezużyteczna – pozostaje niezbędna w komunikacji klinicznej, badaniach czy organizacji leczenia. Oznacza jednak, że kategoria diagnostyczna jest mapą pewnego poziomu organizacji, a nie pełną teorią osoby. W epoce AI szczególnie wyraźnie ujawni się różnica między diagnozowaniem stanu człowieka a modelowaniem systemu, w którym ten człowiek funkcjonuje. Terapia może zatem zmienić swój podstawowy przedmiot: o ile w modelu tradycyjnym terapeuta pomaga pacjentowi zmienić coś wewnątrz siebie, o tyle tutaj obszar interwencji ulega rozszerzeniu.

Systemowe ujęcie sprawstwa i patologii w relacji człowiek-AI

W modelu rozproszonym interwencja może być skierowana jednocześnie do człowieka i jego środowiska: poprzez zmianę sposobu otrzymywania informacji, częstotliwości bodźców czy reguł konsultowania decyzji z AI. Można w ten sposób wpłynąć na dostępność zachowań unikowych, charakter cyfrowej walidacji oraz sposób wykorzystania technologii jako rusztowania.

Mahout staje się wtedy pojęciem systemowym, a nie kolejnym homunkulusem ukrytym w głowie. Nie trzeba szukać w mózgu miejsca, w którym fizycznie by przebywał; jest nim raczej funkcja organizacji sprzężenia zwrotnego między impulsem, refleksją, celem, informacją a środowiskiem. Funkcję tę realizuje czasem sam człowiek, innym razem druga osoba lub instytucja, a coraz częściej uczestniczy w niej algorytm. Najbardziej doniosłą zmianą nie będzie zatem powstanie maszyny, która „myśli jak człowiek", lecz fakt, że możemy otrzymać świat, w którym człowiek myśli inaczej właśnie dlatego, że myśli stale razem z maszyną.

To przesunięcie dotyka samego rdzenia niniejszego artykułu. Klasyczny model Słonia i Jeźdźca skupiał się na problemie kontroli wewnętrznej. Model Mahouta wprowadził relację i uczenie. Z kolei zewnętrzny Mahout dodaje trzeci wymiar: governance regulacji. Nie wystarczy już ustalić, jak uspokoić Słonia i skorygować Jeźdźca.

Konieczne jest ustalenie, kto buduje system im pomagający, jakie dane o nich gromadzi, jakie wartości optymalizuje i czy jego obecność zwiększa przyszłą zdolność człowieka do samodzielnego kierowania własnym życiem. W tej perspektywie podstawowym kryterium dobrej technologii psychologicznej nie powinna być jej zdolność do zastępowania człowieka, lecz umiejętność zwiększania jakości całego układu regulacyjnego bez niejawnego przejmowania jego celu.

Rozróżnienie to stanie się szczególnie istotne w kontekście dotychczasowego paradygmatu psychiatrii. Spór rozgrywano tu najczęściej między dwiema skrajnościami: objaw jako efekt choroby wewnątrz mózgu lub objaw jako normalna reakcja na środowisko. Zewnętrzny Mahout pokazuje, że obie formuły są zbyt ubogie. Zachowanie psychiczne może być produktem sprzężenia między podatnością biologiczną, doświadczeniem, relacją, kulturą, instytucją a aktywnie reagującą technologią.

Wówczas pytanie „gdzie znajduje się zaburzenie?" może okazać się źle postawione. Zaburzenie może dotyczyć organizmu, mechanizmu uczenia lub relacji; może wynikać z niedopasowania organizmu do środowiska bądź być wzmacniane przez instytucję. W nowej epoce część patologicznej dynamiki może powstawać w pętli człowiek-algorytm, nawet gdy żaden pojedynczy element tej pętli nie jest sam w sobie „uszkodzony". Takie ujęcie nie znosi medycyny, lecz zmusza ją do rozszerzenia jednostki analizy.

Kolejny krok Fundacji Dobre Państwo powinien być zatem bardziej wymagający niż samo hasło o „przełamaniu starego paradygmatu". Należy zestawić model biomedyczny i biopsychospołeczny, kategorialną psychopatologię DSM/ICD, podejścia wymiarowe, RDoC, HiTOP, psychiatrię ewolucyjną, teorię sieci objawów, predictive processing oraz modele systemowe, a następnie zapytać, który z nich najlepiej radzi sobie z człowiekiem, którego Mahout może już znajdować się częściowo poza nim. Dopiero takie zderzenie pozwoli uczciwie ocenić, czy stary paradygmat jest jedynie kotwicą, czy wciąż zawiera elementy, bez których nowy model utraciłby diagnostyczną precyzję.

> Ostatecznie największa władza zewnętrznego Mahouta nie tkwi w wydawaniu poleceń, lecz w dyskretnej selekcji doświadczeń, z których uczy się nasz Słoń i Jeździec. Czy w świecie, gdzie algorytmy stają się naszymi głównymi regulatorami uwagi i emocji, wciąż

> będziemy autorami własnych decyzji, czy jedynie pasażerami w systemie, który optymalizuje nas według cudzych celów? Prawdziwym wyzwaniem nie jest już stworzenie maszyny myślącej jak człowiek, lecz uniknięcie świata, w którym człowiek myśli inaczej tylko dlatego, że zapomniał, jak robić to bez maszyny.

Inspiracje

[1] **"Your Normal Mind"** John Hannan

John Hannan jest australijskim psychologiem, specjalistą od traumy i pedagogiem, który specjalizuje się w psychologii ewolucyjnej, odporności psychicznej i interwencji kryzysowej. Swoją karierę zawodową rozpoczął w turystyce przygodowej, spędzając dwanaście lat jako przewodnik i menedżer raftingu w Azji i na dalekiej północy Queensland. Po trzydziestce zajął się psychologią, koncentrując się na środowiskach o wysokich potrzebach oraz leczeniu zespołu stresu pourazowego, urazów moralnych, lęków i depresji. Hannan spędził ponad dekadę pracując jako psycholog w Royal Flying Doctor Service, zapewniając przywództwo kliniczne, terapię i reagowanie kryzysowe w odległych społecznościach Aborygenów i mieszkańców Cieśniny Torresa na Półwyspie Cape York. Pracował również w australijskim Departamencie Obrony, wykładał statystykę uniwersytecką oraz prowadził szkolenia z zakresu reagowania na incydenty krytyczne i zdrowia psychicznego dla sektora korporacyjnego i przemysłowego.

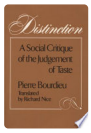
John Hannan *is an Australian psychologist, trauma specialist, and educator who focuses on evolutionary psychology, mental health resilience, and crisis intervention. He began his professional career in adventure tourism, spending twelve years as a white-water rafting guide and manager across Asia and Far North Queensland. In his thirties, he transitioned into psychology, concentrating on high-needs environments and the treatment of post-traumatic stress disorder, moral injury, anxiety, and depression. Hannan spent over a decade working as a psychologist with the Royal Flying Doctor Service, providing clinical leadership, therapy, and emergency crisis response to remote Aboriginal and Torres Strait Islander communities in the Cape York Peninsula. He has also worked within the Australian Department of Defence, taught university statistics, and conducted critical incident response and mental health training for the corporate and industrial sectors.*

Royal Flying Doctor Service

Literatura uzupełniająca

Książki

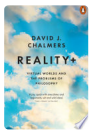
Książki autorów przywoływanych i literatura pokrewna:



[1] "Distinction"

Pierre Bourdieu

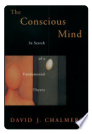
Harvard University Press | ISBN: 9780674212770



[2] "Reality+"

David Chalmers

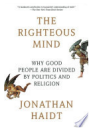
Penguin UK | ISBN: 9780241320723



[3] "The Conscious Mind"

David Chalmers

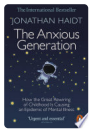
Oxford University Press | ISBN: 9780199839353



[4] "The Righteous Mind"

Jonathan Haidt

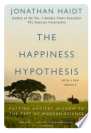
Vintage | ISBN: 9780307907035



[5] "The Anxious Generation"

Jonathan Haidt

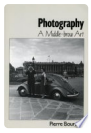
Random House | ISBN: 9781802063288



[6] "The Happiness Hypothesis"

Jonathan Haidt

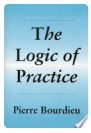
Random House | ISBN: 9781448166688



[7] "Photography"

Pierre Bourdieu

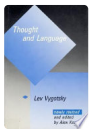
Stanford University Press | ISBN: 9780804726894



[8] "The Logic of Practice"

Pierre Bourdieu

Stanford University Press | ISBN: 9780804720113



[9] "Thought and Language"

Lew Wygotski

MIT Press | ISBN: 9780262720106

Artykuły naukowe

Prace autorów przywoływanych w tekście:

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

Andy Clark

2015

Źródło

[2] "The Extended Mind"

Andy Clark, David J. Chalmers

1998 | Analysis | DOI: 10.1093/analys/58.1.7

Źródło

[3] "Supersizing the Mind"

Andy Clark

2008 | DOI: 10.1093/acprof:oso/9780195333213.001.0001

Źródło

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

Mark Erickson, Andy Clark

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

Źródło

[5] "The Conscious Mind"

David J. Chalmers

1996

Źródło

[6] "Philosophy of Mind : Classical and Contemporary Readings"

David J. Chalmers

2021 | Medical Entomology and Zoology

Źródło

[7] "Perception and the Fall from Eden"

David J. Chalmers

2006 | Oxford University Press eBooks | DOI: 10.1093/acprof:oso/9780199289769.003.0003

Źródło

[8] "Distributed cognition in an airline cockpit"

Edwin L. Hutchins, Tove Klausen

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

Źródło

[9] "The social organization of distributed cognition."

Edwin L. Hutchins

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

Źródło

[10] "The Technology of Team Navigation"

Edwin L. Hutchins

2014 | Psychology Press eBooks | DOI: 10.4324/9781315807645-16

Źródło

[11] "The emotional dog and its rational tail: A social intuitionist approach to moral judgment."

Jonathan Haidt

2001 | Psychological Review | DOI: 10.1037/0033-295x.108.4.814

Źródło

[12] "Social Functions of Emotions at Four Levels of Analysis"

Dacher J. Keltner, Jonathan Haidt

1999 | Cognition & Emotion | DOI: 10.1080/026999399379168

Źródło

[13] "Body, Psyche, and Culture: The Relationship between Disgust and Morality"

Jonathan Haidt, Paul Rozin, Clark Mccauley, Sumio Imada

1997 | Psychology and Developing Societies | DOI: 10.1177/097133369700900105

Źródło

[14] "The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity)."

Paul Rozin, Laura Lowery, Sumio Imada, Jonathan Haidt

1999 | Journal of Personality and Social Psychology | DOI: 10.1037/0022-3514.76.4.574

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: 5c5c28d8-214f-49ba-bf66-3d0769337e45