Wielopoziomowy model korygowalnej regulacji w kontekście książki Your Normal Mind Johna Hannana

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje przejście od dychotomicznej metafory Słonia (emocje) i Jeźdźca (racjonalność) ku zintegrowanemu modelowi korygowalnej regulacji, określanemu jako funkcja Mahouta. Zamiast skupiać się na kontroli nad emocjami, autor proponuje podejście oparte na kalibracji systemu, rozpoznawaniu sygnałów i optymalizacji środowiska. Tekst konfrontuje to ujęcie z neuronauką oraz ramami RDoC, podkreślając znaczenie relacji interpersonalnych i technologii w procesie regulacji. Kluczowym punktem jest rozróżnienie między ewolucyjną logiką mechanizmów obronnych a ich kliniczną dysfunkcjonalnością; autor argumentuje, że stany patologiczne nie są zawsze adaptacyjne, lecz mogą stać się samopodtrzymującymi się błędami systemu, wymagającymi zewnętrznej i wewnętrznej korekty.

The article analyzes the transition from the dichotomous metaphor of the Elephant (emotions) and the Rider (rationality) toward an integrated model of corrected regulation, referred to as the Mahout function. Instead of focusing on controlling emotions, the author proposes an approach based on system calibration, signal recognition, and environmental optimization. The text confronts this perspective with neuroscience and the RDoC framework, emphasizing the importance of interpersonal relationships and technology in the regulatory process. A key point is the distinction between the evolutionary logic of defense mechanisms and their clinical dysfunctionality; the author argues that pathological states are not always adaptive but can become self-sustaining system errors requiring both external and internal correction.

Artykuł stanowi krytyczną rewizję potocznego rozumienia ludzkiej psychiki jako pola walki między racjonalnym umysłem a emocjami. Autor odrzuca dychotomiczny model

„Słonia i Jeźdźca” na rzecz koncepcji korygowalnej regulacji, w której centralną rolę odgrywa funkcja „Mahouta” – wielopoziomowego systemu kalibracji zachowań. Teza główna zakłada, że zdrowie psychiczne i wolność nie polegają na absolutnej kontroli nad impulsami, lecz na zdolności organizmu do aktualizowania reakcji w odpowiedzi na nowe informacje. Analiza wykracza poza biologię, włączając w pętlę regulacyjną relacje społeczne oraz technologię AI, która może pełnić rolę zewnętrznego regulatora. Tekst argumentuje, że współczesna psychopatologia nie powinna być postrzegana jako „zepsucie” mózgu, lecz jako utknięcie systemu w samopodtrzymujących się stanach, które utraciły swoją pierwotną funkcję adaptacyjną. W tym ujęciu sprawstwo człowieka definiowane jest jako zdolność do projektowania własnej architektury wyboru i świadomego delegowania funkcji regulacyjnych przy zachowaniu suwerenności nad celami swojego życia.

SŁOWA KLUCZOWE

model korygowalnej regulacji | Your Normal Mind Johna Hannana | metafora Słonia i Jeźdźca | funkcja Mahouta | współregulacja interpersonalna | systemy dynamiczne w psychologii | RDoC (Research Domain Criteria) | rekursywna pętla regulacyjna | suwerenność epistemiczna | rozproszona architektura poznania | adaptacyjne środowisko cyfrowe | inhibitory learning | allostaza | model biospołeczny Linehan

Od dychotomii kontroli do modelu korygowalnej regulacji

Człowiek jako układ regulowany i regulujący: synteza modelu Mahouta w świecie społecznym, biologicznym i algorytmicznym

Punktem wyjścia całego rozumowania była metafora Słonia i Jeźdźca, wykorzystana w materiale źródłowym do zakwestionowania potocznego obrazu człowieka jako istoty przede wszystkim racjonalnej, której emocje są jedynie zakłóceniami świadomego kierowania zachowaniem. Źródło przedstawia Słonia jako potężny, ewolucyjnie starszy system reagujący emocjonalnie i doświadczeniowo, a Jeźdźca jako znacznie młodszy system refleksyjny, który przecenia zakres

własnej kontroli i często racjonalizuje decyzje powstałe poza świadomością. Jako metafora dydaktyczna obraz ten trafnie narusza naiwny racjonalizm. Jako teoria umysłu jest jednak zbyt dychotomiczny.

Współczesna neuronauka nie znajduje dwóch odrębnych podmiotów, z których jeden byłby emocjonalny, a drugi racjonalny. Procesy emocjonalne uczestniczą w wartościowaniu decyzji, poznawcze modulują emocje, uczenie deklaratywne i niedeklaratywne oddziałują na siebie, a świadoma refleksja może zarówno korygować reakcję, jak i dostarczać jej nowych racjonalizacji. Słoń i Jeździec powinny więc pozostać metaforą dwóch klas funkcji, a nie anatomią psychiki.

Właśnie w tym miejscu potrzebny okazał się Mahout. Nie jako trzecia część mózgu ani niewidzialny homunkulus siedzący ponad Słoniem i Jeźdźcem, lecz jako nazwa funkcji długoterminowej regulacji. Mahout nie zwycięża organizmu własną siłą; jego przewaga polega na akumulacji wiedzy o konkretnym systemie, rozpoznawaniu regularności, przewidywaniu sytuacji, przygotowywaniu środowiska, uczeniu się z błędów i modyfikowaniu strategii wraz ze zmianą warunków. Metafora ta przesuwa zatem punkt ciężkości z problemu kontroli na problem kalibracji. Pytanie nie brzmi już, czy refleksyjny Jeździec jest wystarczająco silny, aby podporządkować emocjonalnego Słonia, lecz czy cały system potrafi rozpoznawać własne sygnały, odróżniać zagrożenie od fałszywego alarmu, uczyć się z doświadczenia, reagować na informację korygującą i organizować otoczenie w sposób zwiększający długoterminową zdolność działania.

Tak sformułowany Mahout nie jest konkurencyjną teorią wobec neuronauki, psychiatrii ewolucyjnej, teorii systemów dynamicznych, RDoC czy psychologii rozwojowej i społecznej. Jest raczej modelem integrującym poziomy opisu, którego wartość zależy od tego, czy pozwala zestawić procesy opisywane przez różne dyscypliny bez ich arbitralnego ujednolicania. RDoC dobrze pokazuje, dlaczego pojedyncza etykieta diagnostyczna nie musi być podstawową jednostką naukowego wyjaśnienia: bada systemy negatywnej i pozytywnej walencji, procesy poznawcze, społeczne, pobudzeniowe i sensomotoryczne na wielu poziomach analizy, umieszczając je w kontekście rozwoju i środowiska. Sam NIMH podkreśla, że RDoC nie jest alternatywną instrukcją diagnostyczną, lecz ramą badawczą służącą analizowaniu funkcji przebiegających od zakresu normalnego do zaburzonego.

Model Mahouta idzie w podobnym kierunku, lecz poszerza jednostkę analizy o relacje interpersonalne, instytucje i aktywnie reagującą technologię. Kluczowa korekta źródła dotyczy zatem samego pojęcia adaptacji. Zbyt często przechodzi się od stwierdzenia, że określony mechanizm ma zrozumiałą funkcję ewolucyjną, do wniosku, że jego patologiczna postać jest nadal "prawidłowo działającym" systemem. Szczególnie wyraźnie widać to w twierdzeniu, że liczne stany

klasyfikowane jako zaburzenia są w istocie ordered responses zdrowego mózgu reagującego na nienormalne środowisko.

Psychiatria ewolucyjna nie daje podstaw do tak szerokiego wniosku. Randolph Nesse podkreśla, że mechanizmy takie jak lęk czy niski nastrój mogą mieć funkcje obronne, lecz zadaniem perspektywy ewolucyjnej jest również wyjaśnienie, dlaczego naturalna selekcja pozostawiła podatność na stany patologiczne, dlaczego użyteczny mechanizm może działać nadmiernie i dlaczego kompromisy, niedopasowania, stochasticzność i ograniczenia ewolucji generują chorobliwość. Najbardziej adekwatne jest wobec tego rozróżnienie pomiędzy ewolucyjną logiką mechanizmu a kliniczną funkcjonalnością jego aktualnego stanu.

Zaburzenia jako samopodtrzymujące się stany systemu

Ból jest niezbędnym systemem ostrzegania, jednak przewlekły ból neuropatyczny może utrzymywać się bez zachowania proporcjonalności względem aktualnego uszkodzenia. Lęk chroni przed niebezpieczeństwem, lecz nadmierna generalizacja potrafi uczynić neutralne środowisko praktycznie niedostępnym. Podobnie niski nastrój w pewnych okolicznościach może ograniczać kosztowną aktywność, ale ciężka depresja z zahamowaniem psychoruchowym, utratą apetytu i wysokim ryzykiem samobójstwa nie staje się przez to programem adaptacyjnym.

Właściwym pytaniem nie jest zatem to, czy objaw jest adaptacją, lecz: z jakiego mechanizmu wyrasta, jaka była jego pierwotna funkcja, dlaczego system generuje obecnie taki stan oraz czy pozostaje on proporcjonalny i podatny na korektę. Taka perspektywa pozwala zachować wartość zasady czujnika dymnego bez przekształcania jej w uniwersalną teorię psychopatologii. Mechanizmy alarmowe rzeczywiście mogą być ukształtowane tak, aby fałszywy alarm był mniej kosztowny niż przeoczenie realnego zagrożenia.

NIMH definiuje normalną reakcję na ostre zagrożenie jako aktywację defensywnego systemu motywacyjnego służącego ochronie organizmu. Jednak w rzeczywistej psychopatologii koszt fałszywych alarmów może stać się ogromny, a system może wejść w pętlę, w której unikanie uniemożliwia zdobycie informacji korygującej. To, co było rozsądnym asymetrycznym biasem bezpieczeństwa przy pojedynczym zagrożeniu, może przeobrazić się w mechanizm utrzymujący przewlekłe ograniczenie funkcjonowania. Trauma szczególnie wyraźnie pokazuje przejście od logiki obronnej do dynamiki zaburzenia. Po zdarzeniu zagrażającym życiu organizm powinien zwiększyć wagę informacji pozwalających uniknąć powtórzenia tego doświadczenia. Pamięć staje się selektywnie uwrażliwiona na sygnały podobieństwa, uwaga kieruje się ku potencjalnym zagrożeniom, a reakcje autonomiczne przyspieszają przygotowanie do obrony. Problem pojawia się

wtedy, gdy generalizacja obejmuje zbyt szerokie klasy sytuacji, pamięć zagrożenia pozostaje wysoce dostępna mimo radykalnej zmiany kontekstu, a zachowania unikowe redukują lęk krótkoterminowy kosztem długofalowej aktualizacji modelu.

W takim stanie pacjent nie jest ani „zepsutym mózgiem", ani po prostu zdrowym systemem wykonującym właściwe zadanie. Jest dynamicznym układem, którego mechanizmy mają zrozumiałą historię, lecz aktualnie wytwarzają kosztowny i samopodtrzymujący się stan. Teoria systemów dynamicznych dostarcza temu opisowi bardziej formalnego języka. Współczesne koncepcje ujmują zaburzenia psychiczne jako alternatywne stany przyciągające złożonego systemu, do których układ może zostać przesunięty przez stresor, a następnie w nich pozostać wskutek wzajemnego podtrzymywania się procesów. Rezyliencja staje się w tym ujęciu właściwością dynamiki i „głębokości" basenu zdrowego funkcjonowania, a nie statyczną cechą charakteru. Model Mahouta można zatem sformalizować jako problem przesuwania prawdopodobieństwa między stanami oraz zmiany parametrów, które czynią jedne stany bardziej stabilnymi od innych.

Człowiek jako system wykraczający poza determinizm genetyczny

Terapia, sen, relacja, lek, zmiana pracy, regulacja środowiska cyfrowego czy przerwanie przemocy mogą oddziaływać na różne punkty systemu i mimo odmiennych mechanizmów zmieniać całą trajektorię zdarzeń. Z tej perspektywy nie istnieje jedna, uprzywilejowana „prawdziwa przyczyna" psychopatologii; odpowiedź na pytanie o przyczynowość zależy od poziomu analizy i skali czasu. Podatność genetyczna może zwiększać reaktywność, dzieciństwo kształtować oczekiwania społeczne, a konkretne doświadczenie uruchomić epizod. Bezsenność może go podtrzymywać, izolacja ograniczać regenerację, algorytmiczne środowisko informacyjne nasilać dostęp do treści zgodnych z aktualnym lękiem, zaś brak zasobów ekonomicznych uniemożliwiać opuszczenie stresującego otoczenia. Wszystkie te czynniki są istotne przyczynowo, bez konieczności wyłaniania jednego „zwycięzcy". Interdyscyplinarność nie polega tu na dekoracyjnym zestawieniu biologii, psychologii i socjologii, lecz na rekonstrukcji przepływu przyczyn między poszczególnymi poziomami. Kluczowym elementem syntezy jest odejście od metafizyki "genów jako zwierzchnika". Materiał przedstawia Słonia jako system służący przede wszystkim ciągłości linii genetycznej, a człowieka jako czasowego nosiciela interesów genów. Jest to efektowna radykalizacja gene-centred language, jednak nieprecyzyjna jako literalna teoria psychologiczna – geny bowiem nie posiadają interesów, projektów ani reprezentacji przyszłości.

Dobór naturalny może wyjaśniać utrzymywanie mechanizmów zwiększających dostosowanie w określonych warunkach, lecz współczesne zachowanie człowieka jest wynikiem wielopoziomowej interakcji dziedziczenia, rozwoju, uczenia się, kultury i instytucji. Mechanizm ukształtowany przez selekcję ze względu na historyczne konsekwencje reprodukcyjne może dziś służyć celom, których sama selekcja nigdy „nie przewidziała". Człowiek nie musi zatem podporządkowywać norm moralnych temu, co mogło zwiększać inclusive fitness przodków. Ewolucja wyjaśnia zdolność do agresji, ale jej nie legitymizuje. Wyjaśnia nepotyzm, lecz nie ustanawia go normą sprawiedliwości. Wyjaśnia znaczenie statusu, ale nie dowodzi, że jego maksymalizacja jest zdrowym celem życia. To rozróżnienie pomiędzy pochodzeniem mechanizmu a zasadą normatywną ma zasadnicze znaczenie dla całej teorii Mahouta. Mahout nie istnieje po to, by lepiej realizować „program genów". Jest nazwą funkcji, dzięki której organizm o ewolucyjnie ukształtowanych predyspozycjach może – w warunkach kultury, refleksji i instytucji – budować cele wykraczające poza pierwotną historię selekcyjną.

Dotyczy to również statusu i walidacji. Hannan zauważa, że pozycja społeczna, przynależność i ocena innych znaczą dla człowieka znacznie więcej, niż sugerowałby model czysto ekonomicznego homo oeconomicus. Jednocześnie przesadne sprowadzenie statusu do miejsca w "kolejce reprodukcyjnej" nie oddaje współczesnej struktury motywacyjnej człowieka. Status może oznaczać dostęp do zasobów, wpływ, prestiż, autonomię, kompetencję, bezpieczeństwo, możliwość bycia wysłuchanym czy reputację moralną. W społeczeństwach kulturowych istnieje wiele częściowo niezależnych hierarchii, a jednostka może świadomie zrezygnować z wysokiego statusu w jednej domenie na rzecz wartości z innej.

Wolność jako stopniowalna zdolność do samokorekty

Walidacja nie jest prostą dawką statusu. Jej kliniczne znaczenie polega przede wszystkim na uznaniu rzeczywistości doświadczenia bez konieczności automatycznego zatwierdzania wszystkich płynących z niego wniosków. To właśnie pozwoliło w części dotyczącej borderline odróżnić rozumienie genezy zachowania od akceptacji działań destrukcyjnych. W modelu biospołecznym Linehan wrażliwość emocjonalna i inwalidujące środowisko mogą wspólnie przyczyniać się do rozwoju dysregulacji, jednak ani trauma, ani sama inwalidacja nie stanowią uniwersalnej i wystarczającej etiologii BPD. Z tej przyczyny próba zastąpienia borderline konstruktem opartym na potraumatycznej inwalidacji może być inspirującą krytyką pewnych sposobów konceptualizacji, lecz nie może być przedstawiana jako ustalona alternatywa nozologiczna. BPD ujawnia jednak coś istotniejszego dla teorii Mahouta niż sama kwestia nazewnictwa diagnozy.

Pokazuje ona, że regulacja ma charakter relacyjny. Stan jednego człowieka zmienia interpretację zachowania drugiego; ta interpretacja modyfikuje odpowiedź, a odpowiedź wpływa na zachowanie partnera, co zwrotnie staje się dowodem potwierdzającym pierwotne oczekiwanie. Strach przed odrzuceniem może w określonych warunkach prowadzić do kontrolowania partnera, co z kolei zwiększa jego dystans – ten zaś zostaje odczytany jako potwierdzenie przewidywanego porzucenia. W takim układzie diagnoza jednej osoby jest niewystarczająca, gdyż jednostką przyczynową jest częściowo pętla interpersonalna. Teoria Mahouta musi zatem uwzględniać współregulację od najwcześniejszych etapów życia. Niemowlę nie posiada autonomicznej zdolności regulowania wszystkich stanów; znaczna część tego procesu zachodzi poprzez opiekuna. Z czasem funkcje interpersonalne mogą zostać częściowo uwewnętrznione jako nawyki uwagi, techniki samouspokajania, język emocji czy oczekiwania wobec relacji. Dorosły nadal jednak korzysta z innych ludzi jako regulatorów: partnera, przyjaciela, współpracownika, lekarza, terapeuty czy wspólnoty. Samoregulacja nie zastępuje współregulacji, lecz wyrasta z jej historii i pozostaje z nią sprzężona. To właśnie dlatego samowspółczucie nie powinno być redukowane do sympatycznego dodatku do terapii – jest ono szczególną formą uwewnętrznionej relacji regulacyjnej.

Człowiek może nauczyć się reagować na własny błąd globalnym atakiem tożsamościowym albo analizą konkretnego zachowania. W pierwszym przypadku informacja „zrobiłem źle” łatwo przeobraża się w „jestem zły”, co zwiększa motywację do obrony przed tą informacją. W drugim możliwe staje się utrzymanie odpowiedzialności bez globalnego samoponiżenia. Samowspółczucie nie polega zatem na obniżaniu wymagań, lecz na redukcji kosztu psychicznego uczciwego kontaktu z własnym błędem. Związek z wolną wolą, przedstawiany niekiedy jako niemal logiczna konieczność, jest bardziej złożony. Fakt, że zachowanie posiada przyczyny i że człowiek nie wybrał swojej biologii, dzieciństwa ani większości warunków formujących pierwsze dyspozycje, podważa naiwne pojęcie absolutnego samostworzenia. Nie prowadzi jednak automatycznie do wniosku o braku wszelkiego sprawstwa. NIMH ujmuje agency między innymi jako zdolność rozpoznawania siebie jako autora własnych działań i myśli, co pokazuje, że sprawstwo jest przedmiotem badania funkcjonalnego, a nie wyłącznie metafizycznym dogmatem.

W rozwiniętym tutaj modelu wolność ma charakter regulacyjny i stopniowalny. Nie polega na istnieniu podmiotu poza łańcuchem przyczynowym, lecz na zdolności systemu do reprezentowania alternatyw, przewidywania konsekwencji, reagowania na racje, hamowania pewnych reakcji, zmiany środowiska i uczenia się z informacji zwrotnej. Człowiek nie wybiera pierwszej spontanicznej reakcji z absolutnej próżni przyczynowej, ale może uczestniczyć w tworzeniu przyszłych przyczyn własnego zachowania. Może zdecydować, że nie będzie podejmował decyzji pod wpływem alkoholu, zablokować aplikację generującą kompulsywne zachowanie, rozpocząć terapię, poprosić o pomoc lub zaprojektować regułę opóźniającą odpowiedź w konflikcie. Każde z

tych działań ma własne przyczyny, ale właśnie dzięki nim system przyczynowy staje się zdolny do modelowania samego siebie. Taki model pozwala zachować odpowiedzialność bez metafizyki absolutnego autorstwa. Człowiek odpowiada nie dlatego, że sam siebie stworzył, lecz dlatego, że posiada określony stopień zdolności reagowania na normy, rozumienia konsekwencji, uczenia się i naprawiania szkody. Z tego samego powodu odpowiedzialność jest stopniowalna. Dziecko, osoba w ciężkim stanie psychotycznym, człowiek działający pod groźbą i sprawca długo planujący przestępstwo nie posiadają identycznego profilu sprawstwa tylko dlatego, że wszystkie ich zachowania są przyczynowo uwarunkowane.

Prawo, etyka i klinika potrzebują więc oceny konkretnych kompetencji, a nie rozstrzygnięcia metafizycznego sporu o determinizm. Współczucie pozostaje w tym modelu wartością normatywną, a nie automatycznym produktem biologii. Człowiek posiada mechanizmy sprzyjające pokrewieństwu, wzajemności, reputacji i współpracy, ale dysponuje również zdolnością do agresji, dominacji, odwetu i wykluczenia.

Ewolucja nie rozstrzyga, który z tych repertuarów jest moralnie właściwy. Kultura, prawo i etyka tworzą systemy pozwalające przekraczać ograniczenia spontanicznej empatii. Człowiek może respektować prawa osoby, której nie lubi; lekarz może uczciwie leczyć pacjenta, z którym światopoglądowo się nie zgadza; sąd może zagwarantować procedurę osobie budzącej odrazę społeczną. Moralna dobroć nie jest więc maksymalizacją empatii, lecz zdolnością do uwzględniania dobra innych również tam, gdzie spontaniczne emocje nie zapewniają wystarczającej motywacji.

Rekursywna pętla regulacyjna człowieka i technologii

Wraz z wprowadzeniem technologii do środowiska regulacyjnego struktura modelu ulega radykalnej zmianie. Intuicja, że media społecznościowe stanowią nowe środowisko, a nie jedynie scenografią dla dawnego umysłu, była trafnym punktem wyjścia, choć twierdzenia o bezpośrednim przekładzie powiadomień na ewolucyjne alarmy wydają się zbyt dosłowne. Istota zmiany tkwi w tym, że technologia modyfikuje skalę, częstotliwość, widzialność, trwałość oraz selekcję sygnałów społecznych. Człowiek zyskuje ogromne rozszerzenie możliwości komunikacji, ale trafia jednocześnie w środowisko, w którym reputacja, porównania społeczne i walidacja są liczone, archiwizowane i aktualizowane niemal bez przerwy.

Z perspektywy regulacyjnej kluczowe jest przejście od medium pasywnego do adaptacyjnego. Książka nie zmieni treści kolejnego rozdziału w zależności od tego, jak długo zatrzymaliśmy wzrok na poprzednim; system rekomendacyjny natomiast potrafi dostosować następny bodziec do wcześniejszego zachowania użytkownika. Powstaje pętla, w której człowiek uczy się ze środowiska,

a środowisko uczy się człowieka. To właśnie ta rekursywność, a nie sam fakt istnienia ekranu, stanowi właściwy przedmiot analizy psychologicznej.

Generatywna AI wykonuje kolejny krok: nie tylko selekcjonuje treść, ale wytwarza ją w odpowiedzi na konkretnego użytkownika. Potrafi przechowywać kontekst, rekonstruować wcześniejsze decyzje, generować alternatywne interpretacje, przypominać o celach i dostosowywać język do reakcji rozmówcy. W ten sposób zewnętrzne środowisko zaczyna przejmować funkcje dotąd przypisywane metapoznaniu. Mahout może zatem częściowo znaleźć się poza układem Słoń-Jeździec.

Nie oznacza to jednak, że AI automatycznie staje się częścią ludzkiego umysłu. Spór o granice poznania w ramach koncepcji extended mind pozostaje filozoficznie otwarty. Dla modelu funkcjonalnego wystarcza skromniejsza teza: zewnętrzny system może zostać włączony w pętlę regulacyjną, pozostając poza biologicznym organizmem.

Pluralizm regulatorów i suwerenność nad delegowaniem

Kalendarz, partner, terapeuta, procedura, aplikacja czy AI mogą w różnym stopniu wpływać na prawdopodobieństwo zachowania, pamięć, dostępność opcji oraz interpretację sygnałów. Ta możliwość wymusza rozwarstwienie pojęcia Mahouta. Mahout wewnętrzny obejmuje metapoznanie, strategie regulacyjne, planowanie i zdolność do obserwacji własnych reakcji. Mahout interpersonalny odnosi się do współregulacji oraz informacji zwrotnej otrzymywanej od innych ludzi. Mahout instytucjonalny obejmuje normy, procedury, system ochrony zdrowia, szkołę, prawo i pracę — wszystko to, co organizuje warunki zachowania. Z kolei Mahout technologiczny obejmuje narzędzia zdolne do monitorowania, przypominania, rekomendowania, filtrowania i coraz częściej interpretowania.

Nie jest to jednak hierarchia, w której wyższy poziom powinien przejąć kontrolę nad niższym. Istotą modelu jest wzajemna korygowalność regulatorów. Człowiek może mylić się w samoocenie i skorzystać z pomocy terapeuty; terapeuta natomiast może błędnie zinterpretować sytuację, co zostanie skorygowane przez doświadczenie pacjenta. Rodzina bywa w stanie zauważyć wzorzec niewidoczny dla obu stron. Dane podłużne mogą ujawnić regularność nieobecną w pamięci, a AI wykryć niespójność w historii. Równocześnie system AI może wygenerować fałszywą interpretację, rodzina stać się źródłem przemocy, terapeuta posługiwać się błędnym modelem, a sam pacjent uciekać w racjonalizację zachowania. Żadne źródło nie powinno więc otrzymać statusu epistemicznego monarchy.

Taka pluralizacja jest szczególnie istotna, gdyż rozwój AI zwiększa ryzyko koncentracji władzy regulacyjnej. Zewnętrzny Mahout, dysponujący obszerną historią danych, może dostrzegać regularności, których użytkownik nie pamięta. Może przewidywać momenty zwiększonej podatności, dostosowywać argumentację i sugerować konkretną decyzję. Z technicznego punktu widzenia może być doskonałym doradcą, lecz z perspektywy politycznej i etycznej natychmiast pojawia się pytanie: czyj cel optymalizuje?

To właśnie jest kluczowe rozszerzenie pierwotnego problemu Słonia i Jeźdźca. Nie wystarczy już pytać, czy Jeździec kontroluje Słonia; trzeba zapytać, kto kontroluje Mahouta. Jeśli system rekomendacyjny żywi się uwagą użytkownika, jego funkcja celu może być sprzeczna z długoterminowym dobrostanem człowieka. Jeśli system monitorujący pracownika należy do pracodawcy, może zostać zoptymalizowany pod kątem wydajności, a nie autonomii czy kosztu allostatycznego. Gdy narzędzie zdrowotne jest powiązane z konkretnym dostawcą usług, pojawia się ryzyko konfliktu interesów.

Sama personalizacja nie jest zatem wartością samą w sobie. To zdolność techniczna, której znaczenie zależy od celu, transparentności, możliwości odmowy i sposobu rozłożenia korzyści. Przyszłość regulacji AI powinna być zatem analizowana w kategoriach zbliżonych do teorii instytucji. Duża zdolność oddziaływania wymaga ograniczeń, audytu, contestability, prawa wyjścia oraz kontroli nad delegowanymi funkcjami. Autonomia człowieka nie polega na konieczności wykonywania każdej operacji samodzielnie — cywilizacja zawsze opierała się na delegowaniu. Polega ona natomiast na zachowaniu dostatecznej suwerenności nad warunkami delegowania: człowiek powinien mieć świadomość korzystania z regulatora, rozumieć jego przybliżoną funkcję, móc kwestionować wynik w istotnych domenach i wycofać określone pełnomocnictwo. Właśnie dlatego AI może zarówno zwiększać, jak i zmniejszać wolność regulacyjną. Osobie z deficytem funkcji wykonawczych może pomóc w organizacji dnia, zwiększając realny zakres samostanowienia; osobie sparaliżowanej lękiem — w przygotowaniu pierwszej ekspozycji, a osobie starszej — kompensować problemy z pamięcią.

Funkcja narzędzia decyduje o korygowalności systemu

System używany kompulsywnie do uzyskiwania pewności może jednak stać się cyfrowym reassurance seeking; system interpretujący każdą kłótnię może osłabiać zdolność mentalizowania, a zawsze przytakujący rozmówca utrwalać jednostronną narrację. Z perspektywy terapeutycznej kluczowym kryterium nie jest zatem sama obecność AI, lecz jej funkcja w mechanizmie problemu. Ta sama aplikacja może pełnić przeciwstawne role u dwóch różnych osób. Pytanie kliniczne

przyszłości powinno więc przypominać klasyczną analizę funkcjonalną: co poprzedza korzystanie z systemu, czego człowiek szuka, jaki natychmiastowy skutek uzyskuje, co dzieje się później i czy zachowanie to zwiększa, czy raczej ogranicza kontakt z doświadczeniem korygującym. Tak sformułowane pytanie łączy erę AI z klasyczną teorią unikania.

Użytkownik może wykorzystać technologię, by przygotować się do rozmowy i następnie ją przeprowadzić – w tym przypadku zewnętrzny Mahout zwiększa ekspozycję. Może jednak użyć jej po to, by rozmów z ludźmi unikać całkowicie; wtedy technologia staje się mechanizmem podtrzymującym unikanie. Różnica nie tkwi w samym narzędziu, lecz w pozycji narzędzia w pętli zachowania. Podobna zasada odnosi się do farmakoterapii, której pierwotnie przypisywano rolę głównie „znieczulania Słonia". Zakładano nawet, że takie znieczulenie może uniemożliwiać habituację. W modelu systemowym nie ma jednak podstaw do tak jednoznacznego rozstrzygnięcia.

Lek może obniżyć pobudzenie i poprawić jakość snu, co w efekcie umożliwia uczenie się; może również wywołać działania niepożądane lub przynieść niewystarczającą korzyść. Psychoterapia może umożliwić aktualizację doświadczeniową, ale może też okazać się nieskuteczna, zostać źle dobrana lub nie dotrzeć do realnego czynnika środowiskowego. Z kolei zewnętrzna pomoc społeczna może rozwiązać problem, którego nie da się przepracować poznawczo, ponieważ jego źródłem jest trwająca przemoc. Najlepszym następcą opozycji „leczenie objawu versus leczenie przyczyny" jest zatem model interwencji w punkty podtrzymywania dynamiki. W systemie z wieloma sprzężeniami przerwanie pętli w jednym miejscu może zmienić przebieg całości.

Lek poprawiający sen może zmniejszyć reaktywność. Zmiana pracy może obniżyć poziom stresu, dzięki czemu psychoterapia staje się możliwa. Ekspozycja może osłabić unikanie. DBT może zwiększyć repertuar regulacyjny, MBT zredukować pewność błędnych interpretacji relacyjnych, a ACT zwiększyć działanie zgodne z wartościami mimo obecności trudnego afektu.

Zdrowie psychiczne jako korygowalność systemu

Żadna z metod nie posiada monopolu na „głębię". Z tego powodu diagnoza powinna mieć status narzędzia, nie ontologicznego wyroku. DSM i ICD umożliwiają komunikację, badania, organizację świadczeń oraz podejmowanie decyzji terapeutycznych. HiTOP ukazuje wymiarową strukturę psychopatologii, RDoC analizuje systemy funkcjonalne, teoria sieci skupia się na wzajemnych powiązaniach symptomów, dynamical systems na stabilności stanów, a psychiatria ewolucyjna na historii funkcji i podatności. Żaden z tych modeli nie powinien być traktowany jako ostateczna mapa osoby. RDoC samo podkreśla, że procesy psychopatologiczne należy badać w kontekście

rozwoju i środowiska, a także na wielu poziomach – od biologicznych po behawioralne i samoopisowe.

Dlatego „kotwicą starego paradygmatu" nie okazuje się ostatecznie sama klasyfikacja DSM, jak sugeruje materiał. Głębszą kotwicą jest lokalizacjonizm wyjaśnienia: przekonanie, że prawdziwa przyczyna musi znajdować się w jednym miejscu. Raz szuka się jej w neuroprzekaźniku, innym razem w traumie, stylu przywiązania, kapitalizmie, rodzinie, genach, dzieciństwie albo algorytmie. Model Mahouta wymusza rezygnację z tej pokusy. Przyczyny psychopatologii mogą mieć różne skale czasowe, działać na różnych poziomach i wzajemnie modyfikować swoją siłę.

To właśnie odróżnia pluralizm naukowy od eklektyzmu. Eklektyzm twierdzi: „wszystko ma znaczenie". Pluralizm mechanistyczny pyta natomiast: „co ma znaczenie tutaj, przez jaki mechanizm, w jakiej skali czasu, z jaką siłą i co zmieni się po interwencji?". Biopsychospołeczna lista czynników pozbawiona struktury przyczynowej jest słaba; biopsychospołeczny model sprzężeń jest za to naukowo płodny. Na tym tle można ostatecznie uporządkować pojęcie zdrowia. WHO definiuje zdrowie psychiczne jako stan dobrostanu umożliwiający radzenie sobie ze stresami życia, realizowanie zdolności, uczenie się, pracę i uczestniczenie we wspólnocie, uznając jednocześnie jego kontinuum oraz wpływ czynników indywidualnych, rodzinnych, wspólnotowych i strukturalnych. Nasz model rozwija tę definicję w stronę dynamiki.

Zdrowie psychiczne nie jest brakiem silnych emocji, niezmiennością ani zdolnością znoszenia każdego środowiska. Jest przede wszystkim zdolnością systemu do wystarczająco trafnej regulacji w warunkach zmiany. Wymaga to czułości na istotne sygnały, ale nie permanentnego alarmu; stabilności, lecz nie sztywności; zdolności tworzenia więzi, ale nie konieczności podporządkowania się każdej grupie; autonomii, lecz nie izolacji; korzystania z zewnętrznego wsparcia, ale nie bezkrytycznego oddawania mu całego autorstwa. Zdrowy system może odczuwać intensywny lęk w sytuacji realnego niebezpieczeństwa, głęboką żałobę po stracie i moralny gniew wobec krzywdy.

Zdrowie nie wymaga usunięcia tych stanów. Wymaga, aby mogły się one pojawić, pełnić funkcję, podlegać aktualizacji i ustępować lub zmieniać wraz z warunkami. Przeciwieństwem zdrowia nie jest zatem po prostu cierpienie, które bywa nieodłącznym elementem adekwatnego kontaktu z rzeczywistością. Przeciwieństwem zdrowia regulacyjnego są raczej takie konfiguracje, w których system traci zdolność dostosowywania reakcji do zmian: utknięcie w alarmie po ustaniu zagrożenia, w odrętwieniu mimo pojawienia się bezpieczeństwa, w samokrytyce mimo zdobycia informacji o błędzie, w przekonaniu odpornym na dowody, w zachowaniu zabezpieczającym uniemożliwiającym nowe uczenie albo w zależności od regulatora, którego nie można już poddać rewizji. To właśnie dlatego korygowalność jest kategorią nadrzędną wobec kontroli.

Człowiek nie kontroluje powstawania wszystkich impulsów, wspomnień i emocji. Nie kontroluje również całego środowiska ani zachowania innych. Może jednak posiadać w różnym stopniu rozwiniętą zdolność do aktualizowania reakcji po otrzymaniu nowej informacji. W języku cybernetycznym zdrowy regulator nie jest nieomylny; jest zdolny wykryć błąd i zmienić parametry działania. Mahout staje się więc figurą uczenia drugiego rzędu. Nie tylko reaguje na świat, lecz uczy się, jak sam na ten świat reaguje. Może zauważyć, że w określonym stanie każdą niejednoznaczność interpretuje negatywnie; że po braku snu łatwiej eskaluje konflikt; że konkretna platforma uruchamia porównanie statusowe; że w kontakcie z daną osobą wraca dawny wzorzec obronny. Dzięki temu nie musi wygrywać z reakcją w chwili jej największej siły – może wcześniej zmienić warunki.

Sprawstwo jako projektowanie architektury wyboru i ekologii regulacji

W tym miejscu model łączy się z ekonomią behawioralną i teorią precommitment. Racjonalna samoregulacja często polega na wcześniejszym projektowaniu środowiska, a nie na późniejszym wykazywaniu heroicznej siły woli. Człowiek może ustawić automatyczne oszczędzanie, ograniczyć dostęp do bodźca, poprosić kogoś o nadzór, zaplanować przerwę lub wyłączyć powiadomienia. Znajomość własnych podatności pozwala zbudować warunki, w których prawdopodobieństwo pożądanego zachowania wzrasta. Jest to sprawstwo realizowane poprzez architekturę przyszłego wyboru. Zewnętrzny Mahout może tę zdolność radykalnie zwiększyć.

AI potrafi zapamiętać, kiedy użytkownik zazwyczaj rezygnuje z celu, rozpoznawać ten wzorzec i przypominać o wcześniejszym postanowieniu w odpowiednim momencie. Jednak właśnie ta cecha sprawia, że może ono przekształcić precommitment w paternalistyczną kontrolę. Kluczowe pytanie brzmi zatem: czy zewnętrzne ograniczenie zostało ustanowione zgodnie z refleksyjnymi celami osoby i pozostaje przez nią odwoływalne, czy też zostało narzucone przez innego regulatora? Model Mahouta okazuje się więc również modelem podziału władzy nad zachowaniem. W dzieciństwie znaczna część tej władzy spoczywa w rękach opiekuna. W dorosłości rośnie zakres samostanowienia, natomiast w chorobie czy starości niektóre funkcje mogą ponownie wymagać większego wsparcia. Instytucje wyznaczają granice zachowań, a technologie zwiększają możliwości, wprowadzając jednocześnie nowe ograniczenia. Zdrowie regulacyjne nie polega na maksymalizacji jednej formy autonomii, lecz na zdolności do organizowania tego podziału w sposób możliwie zgodny z podmiotowością osoby.

Perspektywa cyklu życia wyjaśnia, dlaczego metafora mahouta jest bogatsza niż obraz Jeźdźca. Jeździec jawi się jako stały kierowca dosiadający niezmiennego zwierzęcia. Tymczasem człowiek zmienia się przez całe życie: dojrzewają funkcje wykonawcze, osobowość wykazuje jednoczesną ciągłość i zmienność, przebudowuje się sieć społeczna, starzenie wpływa na bilans zasobów, a choroba może osłabić zdolności metapoznawcze. Nie istnieje jeden niezmienny wewnętrzny zarządca organizmu. Mahout również się uczy, starzeje i czasem sam potrzebuje mahouta. Prowadzi to do pojęcia ekologii Mahoutów.

Człowiek posiada mechanizmy wewnętrznej regulacji, lecz przez całe życie funkcjonuje w sieci regulatorów interpersonalnych, instytucjonalnych i technologicznych. Wspierające środowisko nie znosi potrzeby samoregulacji, lecz stwarza warunki do jej rozwoju. Środowisko toksyczne może natomiast wykorzystywać naturalne podatności w sposób utrudniający późniejszą samodzielność. Właśnie dlatego transmisji międzypokoleniowej nie należy rozumieć wyłącznie jako przekazywania traumy. Przekazujemy sobie również style regulacji, sposoby naprawy relacji, język emocjonalny, normy współpracy, zdolność przepraszania, edukację oraz zasoby instytucjonalne.

Epigenetyczne mechanizmy biologicznego osadzania stresu pozostają interesującym obszarem badań, jednak twierdzenia o prostym transgeneracyjnym dziedziczeniu traumy poprzez znaczniki epigenetyczne są znacznie mniej pewne niż psychospołeczne mechanizmy transmisji. Model Mahouta nie wymaga biologicznego determinizmu międzypokoleniowego, aby uznać odpowiedzialność za przerwanie szkodliwych pętli. Odpowiedzialność ta nie może zostać przeniesiona wyłącznie na rodzinę.

Zdrowie psychiczne jako dobro publiczne

Warunki materialne, szkoła, bezpieczeństwo, praca rodziców, prawo, dostęp do ochrony zdrowia, dyskryminacja czy środowisko cyfrowe – wszystkie one uczestniczą w rozwoju człowieka. Jeżeli psychologia przyzna środowisku znaczenie tylko po to, by następnie nakazać jednostce „lepiej sobie z nim radzić”, dokona jedynie bardziej wyrafinowanej indywidualizacji problemu społecznego. To samo dotyczy koncepcji rezyliencji: odporność nie powinna oznaczać zdolności do nieskończonego znoszenia niekorzystnych warunków. System może adaptować się do przemocy, chronicznej kontroli czy deprywacji w sposób, który zwiększa szanse na krótkoterminowe przetrwanie, lecz jednocześnie generuje długofalowy koszt. Zdolność funkcjonowania w toksycznym środowisku nie dowodzi zdrowia tego otoczenia ani nie daje instytucjom moralnego prawa do utrzymywania go w niezmienionej postaci.

Zdrowie psychiczne jest zatem częściowo dobrem publicznym. Jednostka może ćwiczyć regulację, jednak jakość środowiska pracy, szkoły, przestrzeni społecznej i przepływu informacji bezpośrednio wpływa na koszt tej regulacji. Organizacja może wymagać od pracownika szczególnej odporności albo zmienić system generujący przewlekłe przeciążenie. Platforma cyfrowa może edukować użytkownika w zakresie higieny korzystania z sieci albo przeprojektować elementy interfejsu. Państwo może finansować terapię lub równocześnie oddziaływać na determinanty społeczne zwiększające ryzyko kryzysów.

Taka perspektywa nie znosi osobistej odpowiedzialności, lecz przesuwa ją z moralnego osądu „powinieneś sobie poradzić" w stronę pytania o realny zakres sprawstwa i dostępne punkty interwencji. Człowiek może być odpowiedzialny za poszukiwanie pomocy, ale jeśli jest ona niedostępna ekonomicznie, jego sprawstwo pozostaje ograniczone. Może odpowiadać za sposób traktowania partnera, choć historia przemocy wyjaśnia, dlaczego wymaga on większej pracy nad regulacją. Wyjaśnienie zwiększa precyzję odpowiedzialności, zamiast ją znosić.

Stanowi to dojrzalszą alternatywę zarówno dla skrajnego indywidualizmu, jak i determinizmu społecznego. Człowiek jest jednocześnie produktem warunków i jednym z warunków dalszego świata. Został ukształtowany przez wcześniejsze przyczyny, lecz poprzez własne działania staje się przyczyną dla innych: dziecka, partnera, współpracownika, instytucji czy przyszłej wersji siebie. Nie jest absolutnie wolny, ale nie jest też tylko miejscem, przez które przyczyny przepływają bez transformacji. Na poziomie filozoficznym koncepcja ta zbliża się do procesowego rozumienia podmiotu. „Ja" nie musi być substancją poza biologią, aby zachowywać ciągłość wystarczającą dla odpowiedzialności, relacji i planowania; jest dynamicznie podtrzymywaną organizacją pamięci, wartości, cielesności, relacji i zdolności działania. Z tej perspektywy Mahout nie jest drugim Ja.

Model Mahouta jako rama analityczna, a nie dualizm

Jest to nazwa dla refleksyjnej i środowiskowej organizacji procesu, dzięki której Ja może pozostawać korygowalne w czasie. Ograniczenie metafory jest tu równie ważne, co jej wartość. Każdy obraz opiekuna i zwierzęcia niesie ryzyko dualizmu: sugeruje istnienie kogoś czystego i racjonalnego oraz czegoś niższego, pierwotnego i wymagającego prowadzenia. W rzeczywistości Mahout i Słoń nie są odrębnymi istotami.

Ten sam mózg generuje alarm, refleksję nad nim, plan ekspozycji i późniejszą zmianę zachowania. Emocje nie są zwierzęcą przeszkodą przeciwstawioną człowieczeństwu; stanowią element systemu, bez którego wartościowanie, motywacja, więź i moralność byłyby niemożliwe. Druga granica metafory wynika z języka opieki. Realny mahout może reprezentować interes człowieka, a nie

słonia, dlatego nie należy romantyzować historycznych praktyk tresury. Analogicznie „dobry regulator" psychiczny nie powinien być definiowany jako ten, który najskuteczniej wymusza zachowanie. Regulacja może stać się przemocą, jeśli cel został narzucony przez zewnętrzną władzę – co jest szczególnie istotne w kontekście rodziny, psychiatrii, pracy i AI. Trzecie ograniczenie dotyczy naukowej specyficzności.

Model Mahouta nie generuje sam w sobie biomarkera, diagnozy ani protokołu leczenia. Jest ramą organizującą pytania, a nie konkurencyjną klasyfikacją psychiatryczną. Jego wartość ocenia się po tym, czy pomaga formułować lepsze hipotezy dotyczące mechanizmów, środowiska, punktów interwencji i rozkładu sprawstwa. Gdyby służył jedynie do przepisania istniejących pojęć na atrakcyjny język Słonia i opiekuna, nie wnosiłby naukowo niczego istotnego. Najbardziej produktywne zastosowanie modelu polega zatem na zadawaniu pytań mechanistycznych: Jaki sygnał system próbuje regulować? Jak powstała aktualna kalibracja? Jaką informację organizm uznaje za wiarygodną? Co podtrzymuje reakcję po ustaniu pierwotnej przyczyny? Które zachowanie redukuje cierpienie krótkoterminowo, ale utrwala problem długoterminowo? Jakie zewnętrzne elementy uczestniczą w pętli? Kto ustala cele tych regulatorów? Jak wygląda możliwość korekty? Gdzie interwencja będzie miała największą dźwignię przy najmniejszym koszcie ubocznym?

W takim ujęciu nawet spory między szkołami psychoterapii stają się bardziej empirycznie uchwytne. Nie musimy rozstrzygać, czy „prawdziwym" mechanizmem jest poznawcza restrukturyzacja, ekspozycja, mentalizacja, akceptacja, regulacja emocjonalna czy relacja terapeutyczna. Możemy pytać, które procesy zmieniają się u konkretnego pacjenta, które mediują poprawę oraz czy różne terapie docierają do wspólnych mechanizmów innymi drogami. Podobnie zamiast sporu „biologia czy środowisko", możemy badać, jak doświadczenie zmienia parametry biologiczne, jak różnice biologiczne wpływają na selekcję środowiska i jak oba poziomy tworzą sprzężenie zwrotne. Zamiast dychotomii „geny czy wychowanie", analizujemy gene-environment correlation, interakcje i ścieżki rozwojowe. Zamiast pytania „człowiek czy AI" – badamy architekturę współsprawstwa. W ten sposób interdyscyplinarność przestaje być zbiorem cytatów z różnych dziedzin, a staje się metodą rekonstrukcji złożonego systemu. Najbardziej wymagającą konsekwencją tej syntezy jest odejście od potrzeby jednego końcowego zwycięzcy teoretycznego.

Ewolucyjna psychiatria tłumaczy, skąd biorą się określone mechanizmy i podatności. Neurobiologia opisuje ich implementację. Psychologia uczenia pokazuje, jak zmieniają się one wraz z doświadczeniem. Psychologia rozwojowa rekonstruuje trajektorię ich powstawania. Teoria przywiązania opisuje część społecznej regulacji, natomiast psychologia społeczna analizuje status, normy, walidację i grupy. Socjologia oraz ekonomia instytucjonalna wskazują zaś na strukturę zasobów i władzy.

Model Mahouta jako wielopoziomowa regulacja sprawstwa

Cybernetyka i teoria systemów opisują sprzężenia zwrotne oraz stabilność. Filozofia rozjaśnia pojęcia sprawstwa, odpowiedzialności i autonomii. Prawo wyznacza granice uprawnionego wpływu, natomiast informatyka wnosi modele zdolne do aktywnego uczestnictwa w pętli regulacyjnej. Żadna z tych dyscyplin nie powinna być sprowadzona do roli przypisu; każda z nich opisuje właściwości niewidoczne na innym poziomie. Właśnie dlatego ostateczną wersję Mahouta można określić jako wielopoziomowy model korygowalnej regulacji.

Słoń nie oznacza już „podświadomości", lecz szeroki zbiór procesów automatycznych, afektywnych, interoceptywnych, motywacyjnych i nawykowych. Jeździec nie jest tożsamy z „logiką", lecz reprezentuje procesy refleksyjne, językowe i wykonawcze, zdolne między innymi do modelowania alternatyw i uzasadniania zachowań. Mahout natomiast nie oznacza trzeciego systemu neuronalnego, lecz funkcję długoterminowej kalibracji i organizacji relacji pomiędzy tymi procesami a środowiskiem. Funkcja ta może być częściowo internalizowana, a częściowo zewnętrzna – realizowana przez pamięć autobiograficzną, metapoznanie, nawyki, partnera, terapeutę, instytucje, prawo czy technologię. Zdrowie regulacyjne zależy więc nie od tego, czy Mahout znajduje się „wewnątrz", lecz od jakości pętli: jej podatności na korektę, zgodności celów, bezpieczeństwa, zdolności uczenia, kosztów oraz zachowania wystarczającego sprawstwa osoby. Tym samym AI nie unieważnia modelu człowieka, lecz czyni jego wcześniejsze uproszczenia bardziej widocznymi.

Nigdy nie byliśmy samowystarczającymi Jeźdźcami. Korzystaliśmy z języka, pisma, innych ludzi, instytucji i narzędzi jako integralnych części systemu poznawczego. AI zwiększa jednak intensywność i adaptacyjność tego zewnętrznego wsparcia do poziomu, przy którym jego regulacyjny charakter staje się niemożliwy do ignorowania. Najważniejsza przyszła granica nie będzie zatem przebiegać pomiędzy „naturalnym człowiekiem" a „sztuczną inteligencją", lecz pomiędzy systemami rozszerzającymi ludzkie sprawstwo a tymi, które przejmują je bez wystarczającej kontroli osoby. To rozróżnienie jest bardziej użyteczne niż pytanie o to, czy AI jest dobra lub zła, ponieważ wiąże ocenę z konkretną architekturą relacji.

Przyszła psychiatria nie powinna opierać się na wyborze między paradygmatem medycznym a antymedycznym. Musi pozostać medycyną tam, gdzie niezbędna jest diagnostyka biologiczna, farmakoterapia, hospitalizacja czy ocena ryzyka; psychologią tam, gdzie przedmiotem zmiany są uczenie, interpretacja i zachowanie; nauką społeczną tam, gdzie źródłem problemu są relacje i instytucje; a coraz częściej również nauką o systemach człowiek-technologia. Nie oznacza to

rozmycia specjalizacji, lecz dokładniejsze wyznaczenie ich granic. Ostatecznie źródłowa teza o potrzebie odejścia od języka "naprawiania zepsutych ludzi" może zostać zachowana jedynie po zasadniczej korekcie. Nie każdy pacjent jest „zdrowym człowiekiem reagującym normalnie na nienormalny świat", gdyż realne dysfunkcje biologiczne, psychologiczne i neurodevelopmentalne istnieją. Jednak człowiek nie jest redukowalny do miejsca, w którym wykryto dysfunkcję. Diagnoza opisuje wybrany poziom funkcjonowania, ale nie wyczerpuje podmiotowości, historii, relacji ani zdolności zmiany. Zrozumienie przyczyn nie jest więc przeciwieństwem leczenia – jest jego warunkiem.

Leczenie nie musi oznaczać naprawiania defektu. Może być przeuczeniem, kompensacją, zmianą środowiska, stworzeniem protezy, przerwaniem sprzężenia, odbudową więzi albo ochroną przed realnym zagrożeniem. W niektórych przypadkach najuczciwszym rezultatem nie będzie „wyleczenie" w sensie powrotu do stanu poprzedniego, lecz recovery: odzyskanie wystarczającego sprawstwa i możliwości prowadzenia znaczącego życia przy pozostającej podatności. Pozwala to również precyzyjniej zdefiniować dojrzałość. Nie jest nią stan, w którym Słoń już nie protestuje, ani taki, w którym Mahout zawsze posiada rację.

Odpowiedzialność jako świadomość realnego zakresu wpływu

Dojrzałość polega na zdolności organizmu do utrzymania różnicy między sygnałem a decyzją, emocją a faktycznym sądem, wyjaśnieniem a usprawiedliwieniem, potrzebą walidacji a koniecznością zgody, współczuciem a rezygnacją z granic oraz pomocą zewnętrzną a oddaniem całej zdolności wyboru. Człowiek rozumiany w ten sposób nie zostaje uwolniony od własnej historii, lecz zyskuje możliwość zmiany relacji z nią oraz modyfikacji warunków, w których historia wywiera wpływ. Nie usuwa traumy z przeszłości, ale może zmienić zakres jej generalizacji. Nie wybiera temperamentu, lecz może poznać jego typowe koszty i zasoby. Nie wybiera wszystkich swoich potrzeb statusowych, ale może wybrać środowiska, w których je realizuje. Nie kontroluje pojawienia się każdego impulsu, jednak może kontrolować część warunków, w których impulsy przechodzą w działanie. Nie stworzył języka, kultury ani instytucji, lecz uczestniczy w ich podtrzymywaniu i zmianie. Nie jest w stanie zatrzymać starzenia, ale może modyfikować sposób organizacji wsparcia. Nie powstrzyma istnienia AI, ale może walczyć o taki ustrój technologiczny, w którym AI zwiększa zdolność człowieka do refleksji zamiast zamieniać jego podatności w mechanizm cudzej optymalizacji.

W tej perspektywie pojęcie odpowiedzialności osiąga najwyższy stopień dojrzałości. Odpowiedzialność nie polega na podtrzymywaniu fikcji, że człowiek stworzył wszystkie warunki własnego zachowania, lecz na rozpoznaniu zakresu wpływu, który rzeczywiście posiada w danej chwili, oraz na kształtowaniu warunków mogących ten zakres zwiększyć. Społeczeństwo odpowiedzialne postępuje analogicznie: nie przypisuje jednostce wszystkich skutków warunków, które samo wytwarza, ale jednocześnie nie odbiera jej zdolności do działania pod pretekstem uwarunkowania. Na poziomie klinicznym oznacza to odejście od dwóch destrukcyjnych narracji. Pierwsza mówi pacjentowi: „wszystko zależy od ciebie", druga zaś: „nic od ciebie nie zależy".

Model Mahouta odpowiada: zakres twojego wpływu jest ograniczony, zmienny i realny; możemy badać, gdzie się znajduje i jak go zwiększać. Na poziomie społecznym w ten sam sposób odrzuca się zarówno pogląd, że każde cierpienie jest prywatną awarią jednostki, jak i przekonanie, że jednostka jest wyłącznie bezsilnym produktem struktur. Instytucje tworzą warunki działania, ale ludzie współtworzą instytucje. Relacja ta jest dwukierunkowa, choć władza w niej bywa skrajnie nierówna.

Na poziomie technologicznym odrzuca się zarówno mit neutralnego narzędzia, jak i katastroficzną wizję maszyny automatycznie odbierającej człowieczeństwo. AI jest elementem środowiska o określonej architekturze interesów, danych i kompetencji. Jej wpływ zależy od tego, jak zostanie osadzona w istniejących relacjach władzy, wiedzy i odpowiedzialności.

Mahout jako proces dynamicznej regulacji i korekty

Właśnie dlatego model Mahouta nie prowadzi do ostatecznej harmonii wszystkich teorii. Przeciwnie – wymaga zachowania "produktywnego konfliktu" między nimi. Funkcjonalność ewolucyjna musi być konfrontowana z kliniczną szkodą, biologiczny mechanizm z biograficznym znaczeniem, a subiektywne doświadczenie z danymi zewnętrznymi. Indywidualna odpowiedzialność powinna współistnieć z warunkami społecznymi, autonomia z zależnością, a współczucie z granicami.

Farmakoterapia i psychoterapia nie mogą być traktowane jako konkurenci walczący o monopol, lecz jako interwencje działające na różnych parametrach systemu. Podobnie ludzki osąd i rekomendacja AI powinny ze sobą koegzystować, przy czym żadna ze stron nie otrzymuje "przywileju nieomylności". Nauka dojrzewa właśnie poprzez utrzymywanie takich napięć wystarczająco długo, by rozstrzygać je empirycznie tam, gdzie jest to możliwe, oraz jawnie przyznawać, gdzie pozostają one sporem normatywnym lub filozoficznym.

Z tej perspektywy pytanie o to, dlaczego zamiast Słonia i Jeźdźca warto rozważyć figurę Mahouta, otrzymuje wreszcie odpowiedź. Nie chodzi o to, że Mahout jest trzecią, „lepszą" częścią umysłu, że człowiek powinien skuteczniej kontrolować emocje, czy że starożytna tradycja opieki nad słoniem dostarcza gotowej psychologii. Mahout jest użyteczny, ponieważ wprowadza czas, relację, uczenie się, indywidualną historię, środowisko oraz odpowiedzialność za sposób regulacji. Jeździec pozostaje figurą decyzji; Mahout jest figurą procesu. Proces zaś najlepiej oddaje to, co współczesna nauka mówi o człowieku: organizm nieustannie przewiduje, uczy się, pamięta i zapomina, reguluje i jest regulowany, modyfikuje środowisko i sam ulega jego modyfikacjom. Granice między tymi operacjami są przepuszczalne, a rozwój nowych technologii tę przepuszczalność jedynie zwiększa.

Ostateczny model nie powinien więc przedstawiać osoby jako Słonia prowadzonego przez jednego opiekuna. Trafniejszy jest obraz systemu wielu regulatorów, którego centralnym przedmiotem troski pozostaje konkretna osoba. Jej ciało, biografia i pierwszoosobowe doświadczenie nie mogą zostać zastąpione przez statystykę populacyjną, instytucję ani AI. Jednocześnie człowiek nigdy nie funkcjonuje w izolacji od relacji, kultury, prawa i technologii. Naukowa wartość koncepcji Mahouta kończy się tam, gdzie zaczyna ona udawać pełną ontologię umysłu. Zaczyna się natomiast tam, gdzie pomaga pamiętać, że proces regulacji posiada historię, kontekst, cele, ograniczenia i możliwość korekty.

W takim ujęciu najlepszym Mahoutem nie jest ten, który nigdy się nie myli, lecz system zdolny rozpoznać własny błąd. Nie jest nim system usuwający każdą nieprzyjemną emocję, lecz taki, który odróżnia emocję informacyjną od reakcji nadmiernie utrwalonej. Nie jest nim system całkowicie niezależny, lecz potrafiący korzystać z pomocy bez utraty zdolności kwestionowania pomagającego. Nie jest nim system pozbawiony przeszłości, lecz taki, którego "przeszłość nie ma monopolu na przyszłość". Te zdania nie są jednak sentencjami samopomocowymi.

Wolność jako korygowalność systemu regulacji

Są to skróty szerszych tez naukowych rozwijanych w całym artykule: o plastyczności i stabilności, o extinction i inhibitory learning, o allostazie, współregulacji, dynamicznych atraktorach, sprawstwie, recovery oraz rozproszonej architekturze poznania. Najważniejszą konsekwencją jest jednak zmiana celu.

Człowiek nie musi pokonać własnego umysłu. Powinien raczej coraz lepiej poznać warunki, w których ten działa, zrozumieć mechanizmy nim kierujące i rozpoznać ich historyczną funkcję oraz aktualne koszty. Chodzi o budowanie takich wewnętrznych i zewnętrznych systemów korekty, które zwiększają prawdopodobieństwo uczenia się zamiast utrwalania błędu. To również

najbezpieczniejszy sposób myślenia o przyszłości AI. Zadaniem inteligentnego systemu nie powinno być stworzenie człowieka, który nie musi już myśleć, podobnie jak celem terapii nie jest pacjent, który zawsze potrzebuje terapeuty. Najbardziej wartościowy zewnętrzny Mahout to taki, który zwiększa przestrzeń rozumienia, ujawnia alternatywy, poprawia pamięć systemu i wskazuje niespójności, pomagając aktualizować model, lecz pozostawiając człowiekowi możliwie szeroki zakres świadomego uczestnictwa w określaniu własnych celów. Jeżeli natomiast AI będzie znała Słonia coraz lepiej, a człowiek coraz mniej rozumiał Mahouta, otrzymamy wynik odwrotny: historycznie bezprecedensową zdolność regulowania zachowania połączoną ze słabnącą zdolnością regulowanego do rozumienia warunków własnej regulacji. Nie byłby to problem „buntu maszyn", lecz znacznie subtelniejszy problem utraty epistemicznej i regulacyjnej suwerenności.

Dlatego najważniejszym wymogiem przyszłego środowiska człowieka nie jest maksymalna inteligencja jego narzędzi, lecz korygowalność całego układu. Korygowalność biologicznego modelu zagrożenia przez nowe doświadczenie. Korygowalność własnego przekonania przez dowody. Korygowalność relacji przez informację zwrotną. Korygowalność diagnozy przez rozwój wiedzy. Korygowalność terapii przez wynik pacjenta. Korygowalność instytucji przez procedury i prawa. Korygowalność algorytmu przez nadzór oraz interes człowieka. Korygowalność Mahouta przez Słonia, Jeźdźca i innych Mahoutów.

Człowiek nie jest projektem ukończonym. Jest systemem zdolnym do rewizji samego siebie, choć nigdy nie dokonuje jej z punktu całkowicie wolnego od wcześniejszych przyczyn. To właśnie zdolność do korekty, a nie metafizyczna niezależność od przyczyn, najlepiej opisuje praktyczną postać ludzkiej wolności.

Model Słonia i Jeźdźca pozwolił dostrzec granice świadomej kontroli. Model Mahouta pozwala pójść dalej: zrozumieć, że rozwój nie polega na wzmacnianiu jednego zwierzchnika nad resztą psychiki, lecz na budowaniu coraz lepiej skalibrowanego systemu regulacji, który pozostaje otwarty na doświadczenie, relację i korektę. W epoce sieci i sztucznej inteligencji system ten po raz pierwszy może w istotnym zakresie wykraczać poza biologiczne granice osoby. Nie unieważnia to człowieka jako podmiotu; przeciwnie – sprawia, że pytanie o warunki jego podmiotowości staje się ważniejsze niż kiedykolwiek. Jeżeli przyszła psychologia, psychiatria, pedagogika, prawo i technologia potraktują tę zmianę poważnie, podstawowym pytaniem przestanie być: „jak sprawić, aby człowiek był lepiej kontrolowany?". Będzie nim pytanie znacznie trudniejsze: jak budować biologiczne, relacyjne, instytucjonalne i technologiczne warunki, w których człowiek pozostaje zdolny do uczenia się, samokorekty, odpowiedzialnego współdziałania i rewizji także tych systemów, które pomagają mu kierować samym sobą.

Metafora Mahouta spełnia swoje zadanie i powinna zostać odłożona. Nie dlatego, że okazała się fałszywa, lecz dlatego, że każda dobra metafora w nauce jest rusztowaniem. Ma umożliwić zobaczenie struktury, by następnie ustąpić miejsca bardziej precyzyjnym pojęciom: regulacji, uczeniu, predykcji, plastyczności, sprzężeniu zwrotnemu, sprawstwu, współregulacji, dynamice systemów, recovery, autonomii i korygowalności. Po jej odłożeniu pozostaje człowiek: biologiczny, historyczny, społeczny, technologicznie rozszerzony, nie w pełni przejrzysty dla samego siebie, podatny na błąd, a zarazem zdolny do tworzenia modeli własnego błędu i zmieniania środowiska, które ten błąd podtrzymuje. Nie umysł, który trzeba pokonać, lecz taki, którego trzeba coraz dokładniej rozumieć – również po to, aby wiedzieć, kiedy nie należy mu wierzyć, kiedy go chronić, kiedy poddać leczeniu, zmienić otoczenie, skorzystać z drugiego człowieka, a kiedy z technologii, która może stać się zewnętrznym Mahoutem. Naukowo najważniejsze pozostaje jednak zastrzeżenie końcowe: żadna z tych dróg nie jest absolutna.

Rozumienie nie zastępuje leczenia, współczucie nie zastępuje odpowiedzialności, autonomia nie zastępuje relacji, adaptacja nie zastępuje norm etycznych, ewolucja nie zastępuje historii indywidualnej, a AI nie zastępuje człowieka tylko dlatego, że potrafi uczestniczyć w części procesów, które wcześniej wymagały człowieka. Dojrzały model musi zachować te różnice. Dopiero wtedy Mahout przestaje być atrakcyjną figurą retoryczną i staje się użytecznym narzędziem interdyscyplinarnego myślenia o psychice: nie jako o polu walki racjonalności z emocją, lecz jako o wielopoziomowym, historycznym i relacyjnym systemie regulacji, którego największą zdolnością nie jest bezbłędność, ale możliwość uczenia się z własnych błędów.

> Czy w świecie, w którym algorytmy coraz lepiej rozumieją nasze instynkty niż my sami, wolność pozostanie zdolnością do samokorekty, czy stanie się jedynie iluzją zarządzaną przez zewnętrznego regulatora? Być może największym wyzwaniem przyszłości nie będzie walka z własnymi emocjami, lecz walka o prawo do popełniania błędów w systemie, który dąży do bezbłędnej optymalizacji. Ostatecznie to właśnie podatność na błąd i możliwość jego naprawy stanowią o tym, co w nas głęboko ludzkie.

Inspiracje

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

John Hannan jest australijskim psychologiem i autorem specjalizującym się w traumie, urazach moralnych i psychologii ewolucyjnej. Zanim po trzydziestce zajął się psychologią, spędził dwanaście lat w turystyce przygodowej jako międzynarodowy przewodnik rzeczny i kierownik operacyjny w Azji i na dalekiej północy Queensland. Następnie Hannan poświęcił swoją karierę kliniczną środowiskom o wysokich potrzebach, w szczególności przez ponad dekadę pracując w Royal Flying Doctor Service. Na tym stanowisku prowadził terapię psychologiczną, interwencje kryzysowe i inicjatywy reagowania na samobójstwa w społecznościach odległych społeczności rdzennych na Półwyspie Cape York. Pracował również w australijskim Departamencie Obrony, wykładał statystykę na poziomie uniwersyteckim i reagował na incydenty krytyczne w sektorze przemysłowym i korporacyjnym. W 2026 roku Hannan napisał książkę „Your Normal Mind", która analizuje ludzkie cierpienie psychiczne z perspektywy biologii ewolucyjnej.

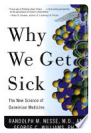
John Hannan *is an Australian psychologist and author specializing in trauma, moral injury, and evolutionary psychology. Before transitioning to psychology in his thirties, he spent twelve years in adventure tourism as an international river guide and operations manager across Asia and Far North Queensland. Hannan subsequently dedicated his clinical career to high-needs environments, notably serving for more than a decade with the Royal Flying Doctor Service. In this role, he delivered psychological therapy, crisis intervention, and community suicide response initiatives across remote Indigenous communities in the Cape York Peninsula. He has also worked within the Australian Department of Defence, taught statistics at the university level, and provided critical incident response for industrial and corporate sectors. In 2026, Hannan authored Your Normal Mind, which examines human psychological suffering through the lens of evolutionary biology.*

Royal Flying Doctor Service

Literatura uzupełniająca

Książki

Książki autorów przywoływanych i literatura pokrewna:



[1] "Why We Get Sick"

Randolph Nesse

Vintage | ISBN: 9780307816009

Artykuły naukowe

Prace autorów przywoływanych w tekście:

[1] "Prevalence and Distribution of Major Depressive Disorder in African Americans, Caribbean Blacks, and Non-Hispanic Whites"

David R. Williams, Hector M. González, Harold W. Neighbors, Randolph M. Nesse, Jamie M. Abelson

2007 | Archives of General Psychiatry | DOI: 10.1001/archpsyc.64.3.305

Źródło

[2] "The National Survey of American Life: a study of racial, ethnic and cultural influences on mental disorders and mental health"

James S. Jackson, Myriam E. Torres, Cleopatra Howard Caldwell, Harold W. Neighbors, Randolph M. Nesse

2004 | International Journal of Methods in Psychiatric Research | DOI: 10.1002/mpr.177

Źródło

[3] "Resilience to loss and chronic grief: A prospective study from preloss to 18-months postloss."

George A. Bonanno, Camille B. Wortman, Darrin R. Lehman, Roger G. Tweed, Michelle Haring, Randolph M. Nesse

2002 | Journal of Personality and Social Psychology | DOI: 10.1037/0022-3514.83.5.1150

Źródło

[4] "What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis"

Eiko I. Fried, Sacha Epskamp, Randolph M. Nesse, Francis Tuerlinckx, Denny Borsboom

2015 | Journal of Affective Disorders | DOI: 10.1016/j.jad.2015.09.005

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: 82b5e7e4-4161-4d7c-b874-1415e8d657a4