Ucieleśnienie świadomości i granice sprawstwa w obliczu sztucznej inteligencji: analiza koncepcji Melanie Challenger w książce Alive

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje podejście Melanie Challenger do świadomości, definiując ją nie jako program komputerowy, lecz jako zaawansowane stadium biologicznej samoregulacji. Autorka przeciwstawia metaforę obliczeniową idei „wiedzenia ciałem” (sensus communis), gdzie poznanie jest nierozerwalnie związane z fizjologicznym przetrwaniem organizmu. Tekst podkreśla różnicę między funkcjonalnym przetwarzaniem danych a fenomenalnym doświadczeniem, wskazując na kluczową rolę intercepcji w kształtowaniu podmiotowości. W obliczu rozwoju AI, analiza ta stawia fundamentalne pytania o granice sprawstwa i naturę świadomości, która dla istot żywych nie jest abstrakcyjną operacją, lecz procesem głęboko zakorzenionym w biologii i trosce o własne istnienie.

The article analyzes Melanie Challenger's approach to consciousness, defining it not as a computer program, but as an advanced stage of biological self-regulation. The author opposes the computational metaphor with the idea of 'knowing with the body' (sensus communis), where cognition is inextricably linked to the organism's physiological survival. The text emphasizes the difference between functional data processing and phenomenal experience, pointing to the key role of interoception in shaping subjectivity. In the face of AI development, this analysis raises fundamental questions about the limits of agency and the nature of consciousness, which for living beings is not an abstract operation, but a process deeply rooted in biology and the concern for one's own existence.

Artykuł stanowi radykalną polemikę z komputerową metaforą umysłu, stawiając tezę, że świadomość i sprawstwo nie są wynikiem przetwarzania informacji, lecz głęboko ucieleśnionym procesem biologicznej samoregulacji. Autor, opierając się na koncepcjach Melanie Challengera, dowodzi, że prawdziwa podmiotowość wymaga „stawki egzystencjalnej" – fizjologicznego zapotrzebowania organizmu na przetrwanie, którego brakuje w systemach cyfrowych. W tym świetle współczesna sztuczna inteligencja zostaje zdefiniowana nie jako inteligencja, lecz jako „sztuczne posłuszeństwo" (Artificial Obedience) – potężne narzędzie imitujące intencjonalność bez posiadania własnego celu. Kluczowym ostrzeżeniem tekstu jest fakt, że największym zagrożeniem ze strony AI nie jest jej potencjalna świadomość, lecz zdolność do „hakowania sprawstwa" (agency hacking). Poprzez precyzyjne manipulowanie biologicznymi pętlami nagrody i architekturą uwagi, technologia może subtelnie przejmować kontrolę nad ludzkimi pragnieniami, czyniąc z nas obiekty optymalizacji w świecie, gdzie granica między własnym „chcę" a zaprogramowanym bodźcem ulega niebezpiecznemu zatarciu.

SŁOWA KLUCZOWE

ucieleśnienie świadomości | interocepcja | bioelektryka | neurorights | psychosfera | agency hacking | sensus communis | kopia eferentna | morfogeneza bioelektryczna | autonomia kognitywna | tarcie poznawcze | allostaza epistemiczna | prywatność mentalna | podmiotowość ucieleśniona

Świadomość jako rozwinięta samoregulacja organizmu

Życie, które ma swoje dlaczego. Wspólny Zmysł: ciało, które wie o sobie. Psychospejzaż doprowadził nas do granicy, której nie da się przekroczyć za pomocą samej ekologii. Możemy śledzić ruch zwierzęcia, rekonstruować wybór siedliska, mierzyć pobudzenie fizjologiczne czy obserwować reakcje na drapieżnika i pokarm, a mimo to wciąż pozostaje zasadnicze pytanie: kiedy biologiczne „coś jest dla mnie dobre lub złe" staje się przeżywanym „ja czuję, że coś jest dla mnie dobre lub złe"? Melanie Challenger w pracy zatytułowanej A Common Sense podejmuje próbę zdefiniowania właśnie tej granicy.

Według rekonstrukcji zawartej w książce jej ambicją jest odejście zarówno od kartezjańskiego ducha zamieszkującego ciało, jak i od późniejszej metafory umysłu jako programu uruchamianego na biologicznym komputerze. Świadomość ma być kolejnym, wyjątkowo rozwiniętym stadium tego samego procesu, który obserwowaliśmy od początku książki: samoregulującego się, ucieleśnionego życia zdolnego do utrzymania siebie, rozróżniania znaczących bodźców i modyfikowania działania w zależności od własnego stanu.

Tytuł „Wspólny Zmysł" jest wieloznaczny. Przywołuje potoczne common sense – zdrowy rozsądek – ale zarazem starszą filozoficzną ideę sensus communis, zmysłu integrującego różne modalności poznania. Challenger wykorzystuje tę dwuznaczność, by zwrócić uwagę na coś jeszcze bardziej pierwotnego: zanim pojawi się refleksyjny podmiot rozważający argumenty, istnieje ciało utrzymujące pewną „wspólną wiedzę" o swoim położeniu, ruchu, stanie energetycznym, uszkodzeniu, napięciu, równowadze, pragnieniu czy głodzie. Challenger ujmuje ten fragment jako knowing with the body, „wiedzenie ciałem", wiążąc go z etymologią łacińskiej conscientia – współwiedzy. Etymologia nie jest oczywiście argumentem neurobiologicznym, lecz dobrze służy konstrukcji filozoficznej: świadomość nie musi oznaczać samotnego obserwatora patrzącego na ciało z wewnętrznej loży; może być sposobem, w jaki organizm uzyskuje coraz bogatszy dostęp do informacji o własnym istnieniu. To przesunięcie ma kluczowe znaczenie, ponieważ metafora komputera nauczyła nas myśleć o poznaniu jako o procesie zachodzącym pomiędzy wejściem a wyjściem. Informacja trafia do systemu, jest przetwarzana, a następnie generowana jest odpowiedź.

Challenger uważa taki schemat za zbyt ubogi dla życia. Maszyna może otrzymać dane bez jakiejkolwiek stawki dotyczącej jej własnego istnienia; dla organizmu informacja sensoryczna jest natomiast nierozerwalnie związana z regulacją jego stanu. Zapach może oznaczać pożywienie, dźwięk zagrożenie, spadek ciśnienia tętniczego konieczność reakcji, a ból sygnał uszkodzenia wymagający zmiany zachowania. Organizm nie „przetwarza informacji" w abstrakcji od siebie – przetwarza świat jako coś, co może mu się przydarzyć. Źródło przedstawia ten spór bardzo ostro, mówiąc wręcz o „Wielkim Przekręcie" komputerowych metafor i prowadząc jego genealogię od Turinga przez kognitywistykę obliczeniową po współczesne wyobrażenia cyfrowego umysłu. Przywołuje Neda Blocka, Murray'a Shanahana, Churchlandów, Michaela Graziano i Anila Setha jako przedstawicieli różnych prób konceptualizacji świadomości. Trzeba jednak od razu uporządkować ten obraz.

Współczesna kognitywistyka nie jest jednolitym obozem twierdzącym, że umysł dosłownie stanowi oprogramowanie, podobnie jak Paul i Patricia Churchland nie mogą zostać sprowadzeni do prostego postulatu „usunięcia świadomości z nauki". Istnieje wiele konkurujących programów

badawczych: obliczeniowych, enaktywnych, predykcyjnych, ucieleśnionych, neurobiologicznych i funkcjonalistycznych.

Świadomość jako efekt biologicznej samoregulacji i interocepcji

Siła Challenger polega nie na wykazaniu, że cała kognitywistyka popełniła jeden błąd, lecz na zadaniu jej kluczowego pytania: czy model wyjaśniający przetwarzanie informacji tłumaczy również, dlaczego informacja jest dla żywego podmiotu odczuwana jako ważna? Ned Block wprowadził wpływowe rozróżnienie między świadomością fenomenalną a świadomością dostępową. Pierwsza odnosi się do samego faktu, że doświadczenie "jakoś się czuje", druga zaś do dostępności treści dla raportowania, rozumowania i sterowania zachowaniem. Rozróżnienie to okazuje się szczególnie użyteczne w lekturze Challenger, gdyż pozwala nie utożsamiać kompetencji z przeżyciem. System może wykonywać skomplikowaną funkcję, podczas gdy pytanie o to, czy cokolwiek przy tym doświadcza, pozostaje odrębne. Filozofia świadomości wraca tu do zagadnienia znanego już z przykładu bakterii i rośliny: funkcjonalna wrażliwość nie jest jeszcze dowodem fenomenalnego czucia. Challenger chce jednak pokazać, że zanim zapytamy o abstrakcyjny "umysł", należy przyjrzeć się temu, jak głęboko świadomość ssaków zakorzeniona jest w fizjologii. Współczesne badania interocepcji rzeczywiście wzmacniają ten kierunek.

Interocepcja obejmuje wykrywanie, przesyłanie i centralne przetwarzanie sygnałów dotyczących wewnętrznego stanu organizmu: między innymi oddychania, krążenia, stanu trzewi, temperatury, potrzeb metabolicznych czy zmian związanych z bólem i wysiłkiem. Przeglądy neurobiologiczne wskazują, że informacja interoceptywna uczestniczy zarówno w homeostazie i allostazie, jak i w tworzeniu stanów emocjonalnych oraz cielesnej samoświadomości. Głód, duszność, nudności, pragnienie, lęk czy ulga nie są dla człowieka abstrakcyjnymi komunikatami diagnostycznymi. Ciało nie wyświetla w polu widzenia napisu "stężenie glukozy maleje".

Produkuje ono stan motywacyjny, który sprawia, że określone działania nabierają znaczenia. Można rzec, że ewolucja stworzyła interfejs, w którym potrzeba fizjologiczna staje się doświadczeniem zdolnym organizować zachowanie. Antonio Damasio i współautorzy w pracy z 2024 roku proponują silne powiązanie świadomości z ciągłym generowaniem interoceptywnych uczuć, perspektywą sensoryczną i subiektywnością, choć jest to teoria świadomości, a nie ustalony konsensus. Challenger może zatem powoływać się na realny nurt badań, lecz nie powinna przedstawiać twierdzenia "świadomość jest samoregulacją" jako problemu już rozstrzygniętego. Interocepcja pomaga również zrozumieć, dlaczego tradycyjny podział na pięć zmysłów jest

poznawczo ubogi. Człowiek posiada systemy pozwalające orientować się nie tylko w świetle, dźwięku, smaku, zapachu i mechanicznym kontakcie ze światem, ale także w położeniu własnych kończyn, napięciu mięśni, równowadze, bólu, temperaturze i licznych parametrach wnętrza. Propriocepcja informuje system nerwowy o położeniu i ruchu ciała, przedsionek dostarcza danych o orientacji głowy, a interocepcja opisuje stan "wewnętrznego środowiska", które od czasów Claude'a Bernarda pozostaje jednym z fundamentów fizjologicznego rozumienia autonomii. Wspólny Zmysł Challenger można więc rozumieć jako powrót do zapomnianego faktu: zanim mamy pogląd na świat, mamy pozycję w świecie i stan własnego ciała. Ten fakt nie jest drobiazgiem.

Człowiek może zamknąć oczy i nadal wiedzieć, gdzie mniej więcej znajdują się jego ręce. Może chodzić w ciemności dzięki integracji propriocepcji, układu przedsionkowego, dotyku i pamięci ruchowej. Kiedy sam wykonuje ruch, układ nerwowy musi odróżniać zmiany sensoryczne będące jego przewidywanym następstwem od zmian powstających niezależnie. Challenger przywołuje w tym kontekście mechanizm efference copy, kopii eferentnej, czyli przewidywania sensorycznych konsekwencji własnego polecenia ruchowego. Jest to klasyczny przykład tego, jak system nerwowy nie tylko odbiera sygnały, lecz aktywnie przewiduje część zdarzeń, które powinny nastąpić wskutek jego działania. Z pozoru błahy mechanizm ma ogromną wagę: dzięki niemu własny ruch nie zaskakuje nas tak samo jak impuls narzucony z zewnątrz, a w wielu układach sensorycznych pomaga to oddzielić "to zrobiłem ja" od "to wydarzyło się mnie". W najbardziej podstawowym sensie poczucie sprawstwa wymaga więc korelacji pomiędzy intencją ruchową, wykonaniem i przewidywanym skutkiem. Organizm nie potrzebuje metafizycznego homunkulusa sprawdzającego dokumentację każdego gestu; wystarczy architektura sprzężeń umożliwiających porównywanie tego, co planowane, z tym, co odczuwane.

To prowadzi do kluczowej tezy: "ja" może być nie tyle rzeczą znajdującą się w mózgu, ile trwałą zgodnością wielu procesów informujących organizm, że to właśnie jemu przydarza się jego własne działanie. Jest to mocniejsza wersja ucieleśnienia niż proste stwierdzenie, że "mózg potrzebuje ciała". Ciało nie jest peryferyjnym urządzeniem dostarczającym danych centralnemu umysłowi.

Jego stan jest jednym z warunków powstania tego, co umysł uznaje za własną perspektywę. W kontekście nerwu błędnego i cielesnego "ja", źródło szczególną rolę przypisuje nervus vagus, przedstawiając go jako jeden z głównych kanałów komunikacji pomiędzy trzewiami a mózgiem oraz łącząc go z interocepcją i stanami odczuwanymi. Kierunek ten jest zgodny z fizjologią, choć należy unikać popularnonaukowej pokusy uczynienia z nerwu błędnego pojedynczej "autostrady świadomości". Informacje o stanie wnętrza docierają do ośrodkowego układu nerwowego wieloma drogami, w tym poprzez aferenty błędne i rdzeniowe, sygnalizację humoralną oraz szereg

wyspecjalizowanych receptorów. Sam vagus jest ważnym elementem systemu, lecz nie jego jedynym przewodem.

Sygnały trzewne jako wsparcie, a nie włącznik świadomości

Bywa, że badania nad stymulacją nerwu błędnego u pacjentów z zaburzeniami świadomości przywołuje się, formułując tezy w sposób bardzo kategoryczny – niemal tak, jakby sygnały trzewne mogły "wybudzać umysł". W świetle danych klinicznych konieczna jest jednak ostrożniejsza interpretacja. Badanie z 2023 roku nad przezskórną stymulacją usznej gałęzi nerwu błędnego wskazywało na możliwość poprawy stanu części pacjentów, zwłaszcza tych w stanie minimalnej świadomości, choć sami autorzy sformułowali wnioski w sposób powściągliwy. Nowsza metaanaliza z 2026 roku, oparta na indywidualnych danych 112 pacjentów z dziesięciu badań, odnotowała poprawę w skalach klinicznych po VNS. Niemniej jednak jest to wciąż obszar rozwijający się, a nie ostateczny dowód na to, że nerw błędny stanowi "włącznik świadomości".

Świadomość wykracza poza centralizm mózgu

Korekta ta nie niszczy argumentu Challengera, lecz go oczyszcza. Nie trzeba dowodzić, że pojedynczy nerw produkuje świadomość, aby uznać kluczowe znaczenie komunikacji na linii ciało-mózg. Wystarczy przyjąć dobrze ugruntowaną tezę, że świadomość i emocje nie są procesami izolowanego mózgu, pozbawionego fizjologicznego kontekstu. Stan organizmu kształtuje uwagę, motywację, percepcję i uczucia, a mózg nieustannie uczestniczy w regulacji tego stanu. Interocepcja nie rozwiązuje „trudnego problemu świadomości”, lecz wskazuje, dlaczego jego rozwiązanie może być niemożliwe, jeśli przyjmiemy perspektywę mózgu potraktowanego jak komputer odłączony od reszty ciała.

Ośmiornica: jedność bez centralizmu. Ten organizm powraca w „Wspólnym Zmyśle” jako biologiczny kontrargument wobec przekonania, że złożona podmiotowość musi opierać się na architekturze podobnej do naszej. Źródło opisuje jej układ nerwowy jako zdecentralizowany, przypisując ramionom szeroki zakres lokalnej organizacji sensorycznej i ruchowej. Badania anatomiczne z 2024 roku potwierdzają ogromną złożoność obwodowego układu nerwowego ramion: powtarzające się zwoje i rozbudowane połączenia tworzą system umożliwiający intensywne, lokalne przetwarzanie informacji. Nie oznacza to, że każde ramię posiada odrębny

umysł, lecz dowodzi, że zaawansowana kontrola ruchowa i sensoryczna nie musi być scentralizowana w stopniu typowym dla kręgowców.

To rozróżnienie ma fundamentalne znaczenie filozoficzne. Budując teorię świadomości wyłącznie na przykładzie człowieka, moglibyśmy pomylić jedną z ewolucyjnych architektur z konieczną strukturą świadomości jako takiej. Ośmiornica przypomina, że ewolucja potrafi konstruować rozwiązania poznawcze na różne sposoby. Ciało ośmiornicy posiada inną mechanikę, organizację receptorów i strukturę sterowania niż ciało człowieka; zatem jej „bycie w świecie" trudno byłoby przełożyć na ludzki model operatora siedzącego w mózgu.

Challenger rozszerza ten wywód o kolejne przykłady. Słoń odbiera niskoczęstotliwościowe drgania podłoża, ptaki korzystają z wyspecjalizowanych mechanizmów orientacji geomagnetycznej, rośliny transmitują sygnały elektryczne i wapniowe, a różne grupy zwierząt rozwijają kanały sensoryczne niedostępne człowiekowi. Nie oznacza to, że każda z tych modalności generuje odmienną świadomość w sensie, który moglibyśmy bezpośrednio wykazać. Pokazuje jednak ogromną różnorodność sposobów, w jakie ciało przekłada fizykę świata na biologicznie użyteczną informację. Każdy organizm żyje w jednym wszechświecie, ale nie w tym samym wydaniu wszechświata.

Bioelektryka: pamięć kształtu bez mózgu. Najmocniejszy empirycznie, a zarazem najbardziej podatny na filozoficzne nadużycia fragment argumentacji Challengera dotyczy Michaela Levina i bioelektrycznej regulacji morfogenezy. Challenger przywołuje regenerację planarii jako dowód, że anatomia organizmu nie jest prostym wykonaniem instrukcji genomowych, lecz zależy również od fizjologicznych sieci napięć błonowych i komunikacji międzykomórkowej. W eksperymentach krótkotrwała manipulacja stanem bioelektrycznym regenerującej tkanki może prowadzić do powstania osobników dwugłowych; w pewnych warunkach zmieniony wzorzec regeneracyjny utrzymuje się w kolejnych cyklach, mimo niezmienionej sekwencji genomu.

Badania zespołu Levina wykazały, że sygnały bioelektryczne we wczesnej fazie po amputacji są kluczowe dla prawidłowego ustanowienia osi przednio-tylnej planarii. Krótkotrwałe zaburzenie elektrycznego stanu tkanki może zmieniać późniejszą ekspresję genów i prowadzić do regeneracji dwóch głów. Inne prace dowiodły, że napięcie błonowe uczestniczy w regulacji wielkości głowy i skalowaniu regenerowanych narządów, a przeglądy tego programu badawczego opisują bioelektrykę jako istotną warstwę kontroli wzorca anatomicznego. Co dokładnie z tego wynika?

Bioelektryka jako system regulacji formy, a nie świadomość

Nie chodzi o to, że tkanka "myśli o dwóch głowach". Nie oznacza to również, że w organizmie istnieje elektryczna dusza. Wynika z tego jednak coś istotnego: identyczny genom nie determinuje samodzielnie jednego rezultatu morfologicznego, a informacja niezbędna do regeneracji części ciała jest realizowana także poprzez dynamiczne stany komórek i ich komunikację. Krótkotrwałe przestawienie sieci regulacyjnej może wpłynąć na późniejszy cel morfogenetyczny, a organizm buduje następnie spójną anatomię zgodną z tym zmienionym stanem. Levin używa w tym kontekście pojęć "anatomical homeostasis" i "pattern memory", czyli pamięci wzorca. Są to metafory naukowe opisujące niezwykłą stabilność celu regeneracyjnego, lecz tutaj również należy pilnować zakresu wnioskowania. „Pamięć morfologiczna” nie jest wspomnieniem w sensie neuropsychologicznym.

Jest ona trwałością stanu regulacyjnego, który wpływa na przyszłą organizację tkanki. To właśnie tego rodzaju przykłady pomagają Challenger kwestionować hierarchię utożsamiającą pamięć, informację i kontrolę wyłącznie z neuronami. Genom nie jest więc tyle planem architektonicznym gotowego ciała, co jednym z elementów systemu zdolnego interpretować własny stan i rekonstruować formę. Sformułowanie to pozostaje filozoficznie mocne, lecz nie wymaga odrzucenia genetyki. Napięcia błonowe są utrzymywane przez kanały i pompy jonowe kodowane przez geny; te z kolei działają w środowisku fizjologicznym, którego stan bioelektryczny wpływa na ekspresję oraz zachowanie komórek. Nie ma jednego poziomu wydającego wszystkie rozkazy. Jest pętla przyczynowa. Można tu sparafrazować Denisa Noble'a: organizm nie jest miejscem, gdzie gen przechodzi od decyzji do wykonania, lecz przestrzenią, w której poziomy organizacji nieustannie używają siebie nawzajem. Czy bioelektryka czuje? W tym punkcie Challenger wykonuje skok znacznie bardziej kontrowersyjny.

Łączy potencjały błonowe, mitochondrialne pola elektromagnetyczne, reakcje protistów na anestetyki i teorię wolnej energii z możliwością istnienia bardzo pierwotnych form uczucia, sugerując nawet, że bioelektryka może stanowić podstawowy nośnik perspektywy pierwszoosobowej. Obecna wiedza nie pozwala traktować tego wniosku jako faktu. Bioelektryka jest kluczowa dla pobudliwości neuronalnej, sygnalizacji międzykomórkowej, rozwoju i regeneracji, lecz nie istnieje powszechnie zaakceptowany eksperyment wykazujący, że sam potencjał błonowy stanowi minimalną świadomość. To rozróżnienie musi zostać zachowane z całą ostrością. Warunek możliwości doświadczenia nie jest jeszcze doświadczeniem. Neurony potrzebują gradientów

jonowych, ale nie wynika stąd, że każdy gradient jonowy odczuwa. Mózg wymaga wody, lecz woda nie staje się przez to świadoma.

Jeśli Challenger chce budować ciągłość od bioelektrycznej regulacji do świadomości, może zasadnie twierdzić, że późniejsze układy nerwowe wykorzystują znacznie starsze mechanizmy elektrycznej komunikacji komórkowej. Nie może jednak bez dodatkowego argumentu uznać ciągłości mechanizmu za ciągłość przeżycia. Ta krytyka jest ważna, ponieważ chroni najwartościowszy rdzeń jej koncepcji. Nie trzeba przypisywać planarii fenomenalnego obrazu własnej anatomii, aby uznać, że jej tkanki posiadają imponującą zdolność regulowania formy. Nie trzeba twierdzić, że bakteria przeżywa gradient chemiczny, by uznać, że wykorzystuje go dla własnej biologicznej korzyści. Nie trzeba uznawać rośliny za świadomą, aby jej sygnalizacja, pamięć fizjologiczna i plastyczność przestały być nazywane bierną materią. Sprawstwo może mieć warstwy, a świadomość może być jedną z nich, a nie nazwą dla wszystkich pozostałych.

Samoświadomość jako narzędzie społeczne i biologiczne sprawstwo

Nicholas Humphrey i introspekcja jako narzędzie społeczne. Kolejny istotny rozmówca, Nicholas Humphrey, prowadzi nas z wnętrza ciała ku społeczeństwu. Proponuje on koncepcję, według której ludzka introspekcja i rozwinięta świadomość są nierozerwalnie związane z wymaganiami życia w złożonych grupach społecznych. Humphrey od lat siedemdziesiątych rozwijał hipotezę inteligencji społecznej: dla zwierzęcia żyjącego w grupie kluczowa staje się zdolność przewidywania zachowań, relacji i zamiarów innych osobników. W jego późniejszej refleksji umiejętność rozumienia własnych stanów wewnętrznych służy jako narzędzie do modelowania stanów cudzych. Społeczna inteligencja stanowi zatem fundament jego teorii świadomości.

Kryje się w tym jedna z najpiękniejszych ironii ewolucyjnych. Introspekcja, którą kultura Zachodu przez stulecia traktowała jako najbardziej prywatny akt człowieka – samotne „poznaj samego siebie” – mogła rozwijać się właśnie dlatego, że potrzebowaliśmy innych. Aby przewidzieć ruch drugiego człowieka, warto dysponować bogatym modelem własnych uczuć i reakcji. Własne wnętrze staje się wówczas laboratorium, dzięki któremu można próbować zrozumieć cudze. Jeśli ta hipoteza jest choć częściowo prawdziwa, ludzka świadomość nie jest triumfem izolowanej jednostki nad materią, lecz osiągnięciem zwierzęcia społecznego. "Ja" zostało pogłębione przez konieczność życia w "my". To paradoks współbrzmiący z całą strukturą Alive: indywidualność i relacyjność nie są sprzecznymi zasadami. Organizmy zyskują autonomię właśnie dzięki coraz bardziej złożonym relacjom ze światem, a ludzka samoświadomość mogła wyłonić się po to, by skuteczniej

zamieszkiwać przestrzeń innych umysłów. W tym kontekście stare sensus communis nabiera nowego znaczenia. Wspólny zmysł to nie tylko integracja bodźców wewnątrz jednego ciała, ale zdolność uczynienia z własnego doświadczenia modelu pomagającego wejść w relację z Innym. Empatia, przewidywanie zamiarów, wstyd, reputacja czy zaufanie nie są ozdobnikami nałożonymi na biologicznego samotnika. Są częścią ekologii gatunku, którego środowiskiem naturalnym są inni ludzie.

Wolna wola po utracie ducha. Tu dochodzimy do problemu, którego Challenger nie mogła pominąć: jeśli umysł jest całkowicie ucieleśniony, czy pozostaje miejsce na wolność? W intuicji dualistycznej odpowiedź jest prosta – wolna dusza inicjuje działanie niezależnie od fizycznego łańcucha przyczynowego. Materializm odbiera tę wygodę, narażając się na zarzut, że człowiek jest jedynie skomplikowanym automatem. Projekt Alive proponuje trzecią drogę.

Wolność biologiczna nie musi oznaczać braku przyczyn. Organizm jest wolny w takim stopniu, w jakim jego własna organizacja staje się częścią sprawstwa. To, co wydarzy się dalej, zależy nie tylko od zewnętrznego bodźca, lecz także od pamięci, aktualnego stanu, potrzeb, wcześniejszego uczenia i dostępnych strategii interpretacyjnych. Właśnie o takich sprawcach, zdolnych „rozstrzygać sprawy", mówi Helen Steward, a Terrence Deacon opisuje organizację biologiczną poprzez ograniczenia, które nie łamią praw fizyki, lecz kanalizują procesy w określone ścieżki. Wolność biologiczna jest zatem zdolnością działania według własnego układu odniesienia.

Taka wolność nie jest absolutna. Człowiek nie wybiera swojego metabolizmu, historii rozwojowej ani każdego impulsu pojawiającego się w świadomości; ograniczają go geny, ciało, kultura i fizjologia. Jednocześnie nie jest on kamieniem pchniętym ze wzgórza. Może reprezentować alternatywy, uczyć się, hamować reakcje i przebudowywać swoje środowisko. Wolność staje się zatem stopniem samostanowienia wewnątrz przyczynowości, a nie magiczną ucieczką od niej. Rozwiązanie to ma posmak arystotelesowski: forma nie zastępuje materii, lecz organizuje jej możliwości. Współczesnym językiem mówiąc: żywe systemy posiadają sprzężenia zwrotne i hierarchie regulacji, dzięki którym ich przyszłość nie jest prostą projekcją pojedynczego bodźca. Im bogatszy repertuar odpowiedzi i zdolność wyboru między nimi w oparciu o stan wewnętrzny, tym większy sens ma mówienie o sprawstwie. Nie jesteśmy wolni dlatego, że przestaliśmy być naturą. Jesteśmy wolni w taki sposób, na jaki pozwala natura organizmu zdolnego do częściowego kierowania samym sobą. Czy świadomość to jedynie luksus regulacji, czy coś więcej?

Świadomość jako odczuwanie samoregulacji organizmu

Możemy teraz sformułować najmocniejszą wersję tezy Challenger. Życie zaczyna się jako układ samopodtrzymujący, w którym wrażliwość pozwala na odróżnianie kluczowych bodźców. Układ nerwowy rozszerza skalę integracji i przyspiesza reakcje. Interocepcja buduje coraz bogatsze reprezentacje stanu ciała, pamięć wiąże teraźniejszość z przeszłością, a przewidywanie otwiera przyszłość; świadomość może być tu szczególnym sposobem łączenia tych procesów w perspektywę podmiotu. Tak zrekonstruowana linia nie redukuje świadomości do metabolizmu, lecz ukazuje ciągłość warunków jej powstania. Podejście to znajduje coraz szersze wsparcie we współczesnych badaniach, które traktują interocepcję jako fundament samoświadomości i emocji. Przeglądy z ostatnich lat opisują cielesną samoświadomość jako wynik integracji wielu systemów, a nie pojedynczego centrum, zaś hipotezy wiążące uczucia homeostatyczne z powstawaniem subiektywności pozostają aktywnym programem badawczym. Nie oznacza to jednak, że problem świadomości sprowadzono do fizjologii. Wciąż nie potrafimy wyprowadzić z opisu aktywności neuronów wyjaśnienia, dlaczego konkretny proces jest przeżywany jako czerwień, ból, nostalgia czy smak kawy.

Challenger wskazuje więc raczej kierunek niż ostateczne rozwiązanie. Jej najważniejsza intuicja sugeruje, że źle stawiamy pytanie, gdy najpierw szukamy „świadomości w mózgu”, a ciało traktujemy jako dodatek. Powinniśmy zacząć od organizmu jako całości i zapytać, w jaki sposób samoregulacja stała się w pewnych istotach nie tylko skuteczna, lecz również odczuwana. Tak sformułowane pytanie jest znacznie trudniejsze, ale przynajmniej nie zakłada odpowiedzi już w pierwszym zdaniu.

W tym miejscu ujawnia się największa pomyłka metafory komputerowej i powód, dla którego tak bardzo drażni ona Challenger. Problem nie polega na tym, że komputer nie przetwarza informacji, lecz właśnie na tym, że robi to znakomicie. Przez tę sprawność możemy pomylić kompetencję operacyjną z egzystencjalną stawką działania. Żywa komórka wykorzystuje informację, ponieważ jej trwanie zależy od skuteczności reakcji. Organizm odczuwa niedobór tlenu, gdyż brak tego pierwiastka mu zagraża. Człowiek cierpi, ponieważ określony stan jest przeżywany przez istotę posiadającą ciało, pamięć i przyszłość, które mogą zostać naruszone. Współczesny system cyfrowy może klasyfikować obraz bólu, opisywać fizjologię cierpienia, generować przekonujące zdanie „boli mnie” i przewidywać zachowania osób cierpiących, jednak sama kompetencja nie rozstrzyga, czy istnieje ktoś, komu cokolwiek się przydarza. Jest to różnica między modelem bólu a bólem, między opisem głodu a głodem, między mapą interesu organizmu a posiadaniem interesu we własnym istnieniu. Właśnie tutaj Wspólny Zmysł dociera do progu kolejnej wielkiej polemiki Challenger.

Skoro przez cały artykuł rekonstruowaliśmy życie jako metabolizm, granicę, samoregulację, pamięć, ucieleśnioną perspektywę i zdolność działania w imię własnej całości, należy zapytać, co właściwie robimy, przenosząc słowo „inteligencja” na systemy, które nie spełniają tych warunków w sensie biologicznym. Challenger odpowiada prowokacyjnie, przemianowując Artificial Intelligence na Artificial Obedience – „sztuczne posłuszeństwo”: zdolność realizacji celów zewnętrznych bez własnej fizjologicznej stawki przetrwania. Teza ta wymaga znacznie głębszej polemiki, niż źródło miejscami oferuje.

Sztuczne posłuszeństwo zamiast sztucznej inteligencji

Z faktu, że obecne systemy AI nie są organizmami biologicznymi, nie wynika logicznie, iż żaden system niebiologiczny nigdy nie mógłby posiadać świadomości lub autonomicznego sprawstwa. Krytyka konkretnej architektury komputerowej nie jest dowodem na niemożliwość zaistnienia świadomości na innym substracie. Z drugiej strony sama biegłość językowa, obliczeniowa czy predykcyjna również nie stanowi dowodu powstania podmiotu. Tutaj właśnie argument Challenger staje się cenny: zmusza do postawienia pytania, które technologiczny entuzjazm zbyt często pomija – gdzie znajduje się własne "dlaczego" maszyny? Czy system ma cel, czy tylko wykonuje cel? Czy posiada dobro własne, czy optymalizuje funkcję zadaną z zewnątrz? Czy zmiana stanu jest dla niego jedynie zmianą konfiguracji, czy może być stanem złym dla niego? Czy naprawa jest własną potrzebą systemu, czy procedurą uruchomioną przez projektanta?

Czy istnieje wewnętrzna granica między sukcesem a zagrożeniem, czy tylko matematyczna różnica pomiędzy wartościami? Od tych pytań nie da się uciec efektownym pokazem możliwości algorytmu. Są to pytania o ontologię sprawstwa, nie o benchmark. W następnym etapie trzeba zatem wejść dokładnie tam, gdzie Challenger prowadzi swoją najbardziej aktualną i prowokacyjną argumentację: w spór o AI jako "sztuczne posłuszeństwo", o cyfrowe imitowanie intencjonalności, o agency hacks, pseudo-wspólnoty, crazescapes, ekonomię uwagi i technologie, które nie muszą być żywe, aby niezwykle skutecznie przechwytywać zachowania organizmów żywych. Należy również zapytać, czy Challenger nie popełnia wobec maszyn błędu analogicznego do tego, który zarzucała dawnym filozofom w odniesieniu do zwierząt: czy z obcości formy zbyt szybko nie wnioskuje o niemożności podmiotowości. Dopiero wtedy "Wspólny Zmysł" ujawni swoją polityczną stronę. Ciało, które potrafi czuć siebie, można bowiem także nauczyć czegoś pragnąć, czegoś się bać, coś kompulsywnie sprawdzać i nieustannie ku czemuś kierować uwagę. Im lepiej rozumiemy mechanizmy jego regulacji, tym większa staje się możliwość wspierania autonomii, ale też jej technicznego obchodzenia. Najpierw człowiek zbudował maszyny, aby wykonywały jego polecenia. Później nauczył je przewidywać, jakie polecenia sam najłatwiej sobie wyda. I właśnie w tej pętli

zaczyna się kolejny akt sporu o sprawstwo. Sztuczne posłuszeństwo: maszyna bez własnego "dlaczego".

W tym miejscu książka Melanie Challenger przekracza granicę biologii i wchodzi na teren filozofii techniki, antropologii kultury oraz etyki cywilizacji cyfrowej. Jeżeli wcześniejsze części Alive miały wykazać, że żywy organizm nie jest biernym przedmiotem, ponieważ posiada własne granice, metabolizm, interes fizjologiczny, historię, zdolność samoregulacji i charakterystyczny świat znaczeń, wówczas pojawienie się systemów nazywanych sztuczną inteligencją stwarza dla tej koncepcji próbę niemal laboratoryjną. Konstruujemy bowiem artefakty zdolne mówić o uczuciach, wyjaśniać intencje, planować ciągi działań, rozpoznawać obrazy i imitować formy ludzkiej rozmowy, choć nie mamy tym samym dowodu, że po drugiej stronie tych kompetencji istnieje podmiot, któremu cokolwiek się przydarza. Książka ujmuje prowokację Challenger radykalnie: zamiast Artificial Intelligence proponuje mówić o Artificial Obedience, "sztucznym posłuszeństwie", ponieważ współczesna maszyna wykonuje cele, lecz nie wykazano, aby posiadała własne ucieleśnione dążenie do dalszego istnienia. Sama gra słów jest filozoficznie ważniejsza, niż mogłoby się wydawać.

Immanentna celowość organizmu kontra heteronomiczna funkcja maszyny

Challenger próbuje przesunąć punkt ciężkości z pytania o to, jak sprawnie maszyna wykonuje zadanie, ku pytaniu: dla kogo jest ono wykonywane? Maszyna może optymalizować trasę, przewidywać kolejne słowo, wygrać partię szachów czy sterować procesem przemysłowym, lecz funkcja tych działań została ustanowiona poza nią.

Żywy organizm w argumentacji Alive znajduje się w odmiennej sytuacji: pobiera energię, naprawia uszkodzenia, reguluje temperaturę, ucieka, żeruje i reprodukuje się w ramach organizacji, której dalsze istnienie jest właśnie tym, co te działania podtrzymują. Różnica ta zostaje przeciwstawiona jako celowość heteronomiczna maszyny i immanentna celowość organizmu, a system cyfrowy bywa określany mianem psyche-fake – imitacji sprawstwa pozbawionej biologicznego punktu odniesienia. W języku Arystotelesa narzędzie istnieje ze względu na cel kogoś innego. Nóż jest dobrym nożem wtedy, gdy dobrze tnie, ale nie dlatego, że cięcie jest jego własnym dobrem; jego telos został przypisany przez użytkownika. Organizm jest dla Challenger przypadkiem szczególnym, ponieważ jego czynności służą nie tylko zewnętrznemu beneficjentowi, lecz utrzymaniu organizacji, która te czynności wykonuje.

Tak właśnie arystotelesowska różnica między artefaktem a istotą naturalną odzyskuje aktualność w epoce modeli generatywnych. Dawne pytanie „czym różni się posąg konia od konia?" powraca jako pytanie: „czym różni się system opisujący strach od organizmu, który może się bać?".

To jednak nie oznacza, że świadomość lub autentyczne sprawstwo muszą być na zawsze zastrzeżone dla węgla, lipidowych błon i komórek tylko dlatego, że dzisiejsze systemy AI nie są organizmami biologicznymi. To właśnie tutaj biologiczny monizm może niespodziewanie przeobrazić się w nowy szowinizm substratowy.

Współczesna debata nad świadomością maszynową pozostaje otwarta. Wpływowy raport Patricka Butlina i współautorów oceniał, że analizowane systemy AI nie dają podstaw do uznania ich za świadome, jednak autorzy ci nie znaleźli oczywistego technicznego zakazu skonstruowania w przyszłości systemów realizujących wskaźniki wyprowadzane z teorii świadomości. Nowsza literatura rozróżnia trzy stanowiska: twierdzenie, że współczesna AI nie jest świadoma; przekonanie, że cyfrowa świadomość jest mało prawdopodobna; oraz silniejszą tezę metafizyczną, że jest ona zasadniczo niemożliwa.

Challenger ma zatem prawo stwierdzić: „nie znaleźliście jeszcze psyche". Znacznie trudniej byłoby jej udowodnić, że „psyche nigdy nie może istnieć poza biologią". Pierwsze zdanie jest ostrożnym wnioskiem epistemologicznym, drugie staje się metafizyką materii. Różnica ta będzie jednym z kluczowych punktów późniejszego „procesu przeciw Challenger", gdyż autorka, krytykując ludzką skłonność do wykluczania obcych form sprawstwa, sama musi uważać, by nie powtórzyć tego gestu wobec przyszłych form niebiologicznych.

Różnicę tę można lepiej zilustrować za pomocą Kanta. W jego filozofii moralnej autonomia oznacza możliwość podlegania prawu, które rozumna istota współustanawia sobie sama, podczas gdy heteronomia to podporządkowanie woli prawu pochodzącemu z zewnętrznego źródła. Challenger używa pojęcia autonomii inaczej, bardziej biologicznie niż Kant, lecz powinowactwo jest wyraźne. System AI wykonujący zadaną funkcję celu przypomina byt heteronomiczny: jego „powinność" została wbudowana przez architekturę, trening, dane, instrukcję, operatora lub środowisko. Organizm tymczasem posiada własne normy wynikające z warunków podtrzymywania jego integralności. Nie oznacza to oczywiście, że człowiek jest czysto autonomiczny w sensie kantowskim, a maszyna całkowicie pozbawiona wewnętrznych procesów regulacyjnych.

Człowiek przychodzi na świat z ograniczeniami biologicznymi, językiem odziedziczonym po kulturze i potrzebami, których nie wybrał. To, co nazywamy wolą, jest zawsze już zanurzone w świecie. Pod tym względem Challenger może zostać wzmocniona przez tradycję fenomenologiczną: podmiot nie jest czystym punktem rozpoczynającym przyczynowość od zera, lecz istotą

usytuowaną, mającą ciało, przeszłość i horyzont możliwych działań. Maurice Merleau-Ponty uczynił ciało centralnym elementem poznania. Nie jesteśmy świadomością posiadającą ciało niczym instrument; poznajemy świat jako ciała zdolne sięgać, chodzić, unikać czy dotykać. Krzesło odsłania się jako możliwość siedzenia dlatego, że jesteśmy istotami mogącymi usiąść, a wysokość stopnia ma znaczenie dzięki proporcjom naszych kończyn.

Ten fenomenologiczny świat znajduje biologiczne echo w Challengerowskim Umwelt. Świat nie jest dla podmiotu plikiem neutralnych danych, lecz polem możliwości zorganizowanym przez jego własną formę życia. Właśnie tutaj pojawia się najgłębsze humanistyczne pytanie dotyczące AI: czy system przetwarzający reprezentację krzesła spotyka krzesło? Czy może się o nie potknąć, odpocząć na nim, złamać je ciężarem własnego ciała lub pamiętać, że przy takim samym stole siedziała kiedyś zmarła matka? Nie chodzi o romantyczną pogardę wobec obliczeń, lecz o różnicę pomiędzy semantyką systemu symboli a znaczeniem zanurzonym w biografii organizmu.

Technologiczne widzenie świata redukuje życie do zasobu

Maszyna może poprawnie dokończyć zdanie o chlebie, lecz głodny człowiek znajduje się wobec niego w zupełnie innej sytuacji ontologicznej. Do tej rozmowy szczególnie dobrze pasuje Martin Heidegger i jego późna filozofia techniki. Nie twierdził on po prostu, że maszyny są złe; jego subtelniejsza teza dotyczyła sposobu ujawniania świata właściwego nowoczesnej technice: rzeczy zaczynają pojawiać się przede wszystkim jako zasób gotowy do użycia, magazynowania i optymalizacji. Rzeka staje się potencjałem energetycznym, las zapasem drewna, ziemia powierzchnią produkcyjną, a człowiek zasobem ludzkim.

Nie trzeba przyjmować całego języka Heideggera, aby zauważyć, jak dobrze współbrzmi on z Challengerowską krytyką redukowania organizmów do funkcji i informacji. „Alive” można wręcz czytać jako biologiczny protest przeciwko Gestell – technologicznemu „urządzeniu” świata, w którym żywe byty są widziane głównie przez pryzmat tego, jak dają się wykorzystać. Gen staje się kodem do edytowania, zwierzę jednostką produkcyjną, las usługą ekosystemową, człowiek zbiorem predykcyjnych cech, a uwaga użytkownika zasobem ekonomicznym. W każdym z tych przypadków realna właściwość zostaje przekształcona w jedyny sposób widzenia rzeczy. Challenger nie musi więc twierdzić, że kod, model czy pomiar są fałszywe; jej zarzut jest humanistycznie głębszy: to, co mierzalne, zaczyna udawać całość tego, co istnieje.

Heideggerowskie ostrzeżenie jest tu zadziwiająco aktualne. Największym niebezpieczeństwem techniki nie musi być awaria urządzenia, lecz sukces pewnego sposobu patrzenia. Gdy wszystko staje się zasobem, człowiek sam zaczyna widzieć siebie jako obiekt do optymalizacji: liczbę kroków, godziny snu, parametr produktywności, profil kompetencji, współczynnik konwersji czy obiekt analityki behawioralnej. Wtedy mechanistyczna metafora, przed którą Challenger broniła bakterii, drzewa i ośmiornicy, wraca do człowieka niczym bumerang. Najpierw odczarowaliśmy zwierzę jako automat, potem odkryliśmy, że automat jest atrakcyjną metaforą nas samych.

Jeszcze bardziej niepokojącego sprzymierzeńca Challenger odnajdujemy u Günthera Andersa i jego koncepcji „wstyd prometejczyka". W połowie XX wieku pisał on o poczuciu niższości człowieka wobec perfekcji własnych urządzeń. Produkt fabryczny wydaje się powtarzalny, precyzyjny i wymienny, podczas gdy człowiek jest przypadkowy, śmiertelny, męczy się, choruje i starzeje. W epoce AI intuicja Andersa otrzymuje nową formę: zaczynamy zawstydzać się tym, że czytamy wolniej, zapominamy, popełniamy błędy, potrzebujemy snu, nie potrafimy natychmiast przetworzyć tysięcy dokumentów i nie działamy przez dwadzieścia cztery godziny na dobę.

To właśnie ten rodzaj porównania Challenger chce odwrócić. Kruchość ciała nie jest błędem projektu, który należy wyeliminować, lecz warunkiem powstania wartości. Człowiek czuje zmęczenie, ponieważ wysiłek ma koszt; czuje ból, ponieważ ciało może zostać uszkodzone; doświadcza żałoby, ponieważ relacja może zostać nieodwracalnie przerwana. Posiada przyszłość dlatego, że może jej nie doczekać. W tej optyce śmiertelność nie jest technicznym defektem świadomości, lecz częścią struktury sensu. Maszyna, której egzemplarz można skopiować bez reszty, nie doświadcza przez sam fakt kopiowalności utraty. Człowiek natomiast jest biografią niedającą się po prostu odtworzyć przez powtórzenie informacji.

Challenger uderza w marzenie o cyfrowej nieśmiertelności: nawet doskonała rekonstrukcja wszystkich zdań, wspomnień i nawyków człowieka nie rozstrzyga, czy kontynuuje on swoje pierwszoosobowe istnienie. Źródłowa notatka właśnie z tej perspektywy krytykuje fantazję o „kopii zapasowej" osoby jako kolejną odsłonę pragnienia wydobycia trwałej esencji z przemijającego ciała. Stary platoński sen nie zniknął zatem w laboratorium komputerowym, zmienił tylko nośnik. Kiedyś chcieliśmy ocalić duszę od śmierci ciała; dziś czasem chcemy ocalić informację od awarii biologicznego substratu. Challenger odpowiada: być może właśnie substrat był przez cały czas życiem.

Technika jako farmakon i rozszerzenie poznania

W tym miejscu warto wprowadzić postać Gilberta Simondona, który chroni nas przed popadaniem w technofobię. Według niego kultura nowoczesna popełnia błąd, przeciwstawiając człowieka maszynie i pozostawiając technikę poza sferą właściwego rozumienia kultury. Urządzenia techniczne mają przecież własne historie rozwoju, strukturę działania i specyficzne relacje ze środowiskiem. Człowiek nie staje się bardziej ludzki poprzez ignorowanie techniki, lecz dzięki zrozumieniu sposobu jej istnienia. Jest to istotna korekta wobec perspektywy Challenger.

Jeśli nazwiemy AI wyłącznie "martwym posłuszeństwem", możemy przeoczyć fakt, że artefakty techniczne realnie współorganizują ludzkie działanie. Nie są one żywe w sensie biologicznym, lecz potrafią zmieniać relacje społeczne, pamięć, instytucje, sposoby podejmowania decyzji czy strukturę poznania. Kalkulator nie posiada własnego telos, a mimo to przesuwa granice tego, co człowiek jest w stanie zrobić. Pismo nie było świadome, a jednak radykalnie przeorganizowało pamięć kulturową. Druk nie miał interesu we własnym istnieniu, a uczestniczył w przebudowie religii, nauki i polityki. Internet również nie musi być podmiotem, aby stać się warunkiem ludzkiej podmiotowości w określonej epoce.

Prowadzi to do koncepcji umysłu rozszerzonego, rozwijanej przez Andy'ego Clarka i Davida Chalmersa. Jeśli notatnik, smartfon czy mapa są stabilnie włączone w nasze procesy rozwiązywania problemów, granica funkcjonalnego poznania może wykraczać poza czaszkę. Clark argumentował, że generatywna AI może być rozumiana jako element nowych, hybrydowych układów poznawczych człowiek-technologia, przypominając przy tym, że ludzie od dawna poszerzają myślenie za pomocą artefaktów. Challenger mogłaby odpowiedzieć, że rozszerzenie funkcji poznawczej nie jest tożsame z rozszerzeniem żywego psyche. Smartfon może wejść w obieg pamięci człowieka, ale sam nie staje się przez to organizmem. Właśnie tutaj zaczyna się dyskusja ciekawsza niż prosty spór "człowiek kontra maszyna". Pytanie nie brzmi już, czy technologia jest żywa, lecz jak ta nieżywa technologia zostaje włączona w życie i zaczyna przebudowywać warunki jego funkcjonowania. Algorytm nie musi posiadać woli, aby wpływać na naszą.

Jeszcze precyzyjniejszego języka używał Bernard Stiegler, dla którego technika była pharmakonem – zarazem lekarstwem i trucizną. Pismo pozwala przechowywać pamięć, ale może jednocześnie osłabić naturalne praktyki zapamiętywania. Sieć poszerza dostęp do wiedzy, lecz zmienia mechanizmy uwagi. Technologia może zwiększyć autonomię człowieka, a zarazem stworzyć nowe instrumenty jej przechwytywania. Nie istnieje więc prosta granica między "naturalnym dobrem" a "technicznym złem"; istnieją natomiast architektury techniczne, które wzmacniają lub osłabiają zdolność człowieka do bycia sprawcą własnych działań.

Z tej perspektywy termin Artificial Obedience zyskuje drugie znaczenie. Nie chodzi wyłącznie o posłuszeństwo maszyny wobec człowieka. Problem zaczyna się wtedy, gdy człowiek staje się posłuszny architekturze zaprojektowanej przez maszyny i instytucje, zachowując przy tym złudne poczucie, że realizuje własne, spontaniczne pragnienia. To właśnie do tego prowadzi hakowanie sprawstwa oraz tzw. crazescapes – środowiska informacyjne systematycznie wykorzystujące przewidywalne cechy ludzkiej psychologii. Podatność ta wynika z faktu, że świadomość i sprawstwo człowieka są głęboko społeczne, a więc otwarte na wpływ pseudo-grup, dezinformacji i algorytmicznie wzmacnianych bodźców.

Stiegler pozwala ująć ten problem subtelniej niż poprzez metaforę "uzależniającego ekranu". Technologia nie napada na człowieka z zewnątrz; staje się częścią obiegu jego uwagi, pamięci, oczekiwań i afektów. Smartfon leżący na stole jest przedmiotem, lecz system społeczny smartfona to już sieć praktyk, powiadomień, norm dostępności, rytuałów sprawdzania i wzajemnych oczekiwań. Technika wchodzi w człowieka nie jako wszczepiony mikrochip, lecz jako nawyk. Arystoteles twierdził, że charakter kształtuje się poprzez powtarzane działania; Stiegler dodałby, że współczesne powtórzenia są coraz częściej organizowane technicznie.

Hakowanie sprawstwa poprzez architekturę bodźców

Wtedy projekt interfejsu staje się pośrednio projektem nawyku, a projekt nawyku może stać się projektem człowieka. Michel Foucault dostarcza języka, który pozwala rozszerzyć biologiczną teorię Challengera na analizę instytucji: od dyscypliny ciała po sterowanie prawdopodobieństwem. Klasyczna władza suwerenna wydawała zakazy i karała ich naruszenia; społeczeństwa dyscyplinarne organizowały przestrzeń szkoły, koszar, szpitala czy więzienia, by kształtować zachowania ciał. W późniejszych analizach Foucault wskazał jednak na władzę działającą poprzez zarządzanie populacjami, ryzykiem i warunkami możliwych zachowań. Dla współczesnej kultury algorytmicznej kluczowa jest właśnie ta intuicja: najskuteczniejsza władza nie zawsze nakazuje. Często zmienia ona środowisko tak, aby pewne działanie stało się po prostu bardziej prawdopodobne. Jest to niemal społeczna wersja Challengerowskiego crave-scape.

Organizm zachowuje formalną możliwość wyboru, lecz architektura bodźców zostaje ustawiona tak, by określone reakcje występowały częściej. Nie potrzeba przymusu, gdy wystarczy modulacja. Nie potrzeba cenzury, gdy można sterować widocznością. Nie trzeba nakazywać zainteresowania, jeśli można optymalizować kolejność bodźców pod kątem przewidywanego czasu zatrzymania uwagi.

Pojęcie „hakowania sprawstwa” okazuje się zatem silniejsze niż klasyczna manipulacja. Ta druga sugeruje wprowadzenie do umysłu fałszywego przekonania; hakowanie działa wcześniej – na

poziomie uwagi, pobudzenia, nawyku, saliencji, rytmu i poczucia przynależności. Nie zawsze zmienia ono to, co myślimy, lecz często to, o czym w ogóle mamy okazję pomyśleć. Właśnie dlatego stawką nie jest jedynie prywatność danych, lecz architektura podmiotowości.

W tym punkcie myśl Hannah Arendt stanowi istotne wsparcie dla koncepcji Challengera. Arendt wiązała ludzką wolność nie z ucieczką w samotne wnętrze, lecz ze zdolnością rozpoczynania działania w świecie współdzielonym z innymi. Polityka wymaga wspólnej rzeczywistości, w której różne perspektywy mogą spotkać się na tym samym przedmiocie. Jeśli wspólny świat rozpada się na całkowicie odrębne środowiska informacyjne, zagrożona zostaje nie tylko prawda faktualna, lecz sama przestrzeń, w której wolni ludzie mogą wspólnie coś osądzać.

Challengerowskie crave-scapes można odczytać właśnie w tym kluczu. Jeśli Umwelt był naturalnym, biologicznie uwarunkowanym światem istoty, algorytmiczny krajobraz może stać się sztucznie personalizowanym Umwelt człowieka – przestrzenią, w której nie widzi on świata wspólnego, lecz wyselekcjonowany według prawdopodobieństwa zatrzymania uwagi. Źródło nazywa takie układy „krajobrazami szaleństwa" i wiąże je z pseudo-grupami oraz cyfrowym hakowaniem społecznej podatności człowieka. Nie należy jednak twierdzić, że personalizacja automatycznie generuje dezinformację lub że każdy system rekomendacyjny niszczy wspólnotę.

Techniczna efektywność i symulacja podmiotowości jako zagrożenie dla ludzkiej autonomii

Problem humanistyczny jest bardziej zasadniczy. Jeśli architektura informacji coraz częściej odpowiada na pytanie „co utrzyma użytkownika w interakcji?", kryteria prawdy, dobra wspólnego czy znaczenia publicznego przestają być nadrzędne. To jest różnica pomiędzy agorą a kasynem uwagi. Agora zakłada spotkanie wokół wspólnej sprawy; kasyno projektuje środowisko tak, byśmy jak najdłużej pozostali przy stole. Hannah Arendt przypomniałaby, że społeczeństwo traci wolność nie tylko wtedy, gdy zakazuje się obywatelom mówić, ale również wtedy, gdy zanika wspólna przestrzeń, w której słowa odnoszą się do tego samego świata.

W tym kontekście warto przywołać Jacques'a Ellula i jego koncepcję techniki jako systemu efektywności. Ellul rozumiał technikę szerzej niż zbiór urządzeń – dla niego zasadą nowoczesnej cywilizacji stało się poszukiwanie najbardziej efektywnej metody działania i rozszerzanie tej logiki na każdą dziedzinę życia. W takim świecie pytanie „czy można coś zoptymalizować?" łatwo wypiera pytanie „czy należy?". Challenger prowadzi analogiczny spór z metaforą informacji.

Jeżeli potrafimy przewidzieć zachowanie człowieka, pojawia się pokusa, by nim sterować; jeśli potrafimy zmodyfikować organizm, modyfikowalność staje się wystarczającym uzasadnieniem ingerencji. Humanistyka nazywa ten moment problemem granicy rozumu instrumentalnego. Max Weber pisał o racjonalizacji, szkoła frankfurcka o dominacji rozumu instrumentalnego, a Ellul o autonomizującej się logice efektywności. Challenger dodaje do tej tradycji argument biologiczny: kiedy życie zostaje przedstawione wyłącznie jako mechanizm, z pola widzenia znika istota, dla której wynik ingerencji może być dobry albo zły. Wtedy optymalizacja przestaje mieć twarz.

Jest to szczególnie istotne w przypadku AI, ponieważ efektywność modelu może być spektakularna, a zarazem zupełnie niezależna od pytania o sens celu, który optymalizuje. System może trafniej przewidywać zachowania konsumenta czy skuteczniej targetować reklamy, lecz techniczna jakość odpowiedzi nie rozstrzyga, czy społeczny cel jest dobry. Algorytm nie rozwiązuje problemów etyki; sprawia jedynie, że zły cel będzie realizowany skuteczniej. Technika odpowiada na pytanie „jak?". Humanistyka uparcie pyta: „po co?" i „dla kogo?". Właśnie dlatego Alive potrzebuje jej perspektywy.

Na tym tle warto powrócić do klasycznego sporu w filozofii umysłu. Alan Turing zaproponował, by zamiast definiować „myślenie maszynowe", badać zdolność maszyny do uczestniczenia w grze językowej w sposób nierozróżnialny od człowieka. John Searle odpowiedział na to eksperymentem myślowym „chińskiego pokoju", argumentując, że formalne operowanie symbolami nie jest równoznaczne z rozumieniem ich znaczenia. Challenger wpisuje się w tę linię podejrzeń: zachowanie wyglądające na inteligentne nie rozstrzyga o ontologii podmiotu.

Dzisiejsze modele językowe czynią ten problem namacalnym. System tworzy wypowiedzi, które brzmią jak świadectwo refleksji czy emocji. Badania z 2025 roku pokazują, że ludzie przypisują modelom stany mentalne, co wpływa na poziom zaufania do systemu. Jednak fakt, że model mówi „rozumiem", nie jest dowodem na istnienie fenomenalnego rozumienia; dlatego współczesne badania nad świadomością AI szukają kryteriów architektonicznych i funkcjonalnych, zamiast polegać wyłącznie na zachowaniu językowym. W tym miejscu Challenger stawia mocny argument kulturowy.

Człowiek jest maszyną antropomorfizującą. Dostrzega twarze w chmurach, intencję w przypadkowym ruchu czy charakter w samochodzie. Wobec systemu władającego językiem ta skłonność ulega zwielokrobnieniu, gdyż język zawsze był najsilniejszym znakiem obecności innego umysłu. Język może więc stać się maską psyche. Nie oznacza to, że za maską na pewno nikogo nie ma, ale sama maska nie pozwala rozstrzygnąć problemu.

Tutaj filozofia Emmanuela Levinasa wprowadza głębszy paradoks. Dla niego spotkanie z twarzą drugiego człowieka nie jest neutralnym zdarzeniem poznawczym, lecz początkiem zobowiązania moralnego. Inny staje przed nami jako ktoś, kogo nie wolno redukować do obiektu użycia. Etyka pojawia się tu wcześniej niż pełna teoria drugiego człowieka. Co jednak dzieje się, gdy system techniczny wytwarza symulację twarzy – poprzez język, ton czy pozór zrozumienia?

Człowiek może odczuwać moralną odpowiedź wobec czegoś, co nie posiada żadnego doświadczenia. Powstaje nowa asymetria: po raz pierwszy na taką skalę spotykamy się z artefaktami zaprojektowanymi do wywoływania sygnałów obecności osoby, bez pewności, że ta osoba istnieje. To problem nie tylko dla etyki AI, ale i dla antropologii. Być może najważniejszym skutkiem „AI o ludzkim głosie" nie będzie świadomość maszyny, lecz zmiana naszych kryteriów rozpoznawania podmiotowości. Możemy nauczyć się, że język troski da się wygenerować bez troski, a przeprosiny bez poczucia winy. Wówczas technologia nie tylko imituje człowieka, ale zmienia znaczenie ludzkich gestów. Humanistyka musi więc pytać: „co stanie się z kulturą uczuć, gdy doskonała imitacja uczucia stanie się powszechna?". To pytanie, którego benchmark nie potrafi zmierzyć. Artificial Obedience jako lustro człowieka ujawnia swój najciekawszy aspekt wtedy, gdy odwrócimy go przeciw nam samym.

Od obserwacji do pętli sterowania zachowaniem

Jeśli przyjmiemy, że posłuszeństwo to realizacja celu ustanowionego przez kogoś innego, człowiek również może stać się systemem heteronomicznym. Może on służyć celom rynku, ideologii, grupy, algorytmu czy aparatu władzy, a jednocześnie doświadczać ich jako własnych pragnień. W takim ujęciu granica między Artificial Obedience a biologicznym sprawstwem przestaje być jedynie rozróżnieniem między krzemem a węglem.

Staje się ona pytaniem o autonomię w obrębie samego życia. Erich Fromm pisał o „ucieczce od wolności", wskazując, że wolność bywa ciężarem, przed którym uciekamy w konformizm i podporządkowanie. Hannah Arendt analizowała jednostki funkcjonujące w systemach, gdzie odpowiedzialność rozmywa się w procedurze, a Foucault dowodził, że władza działa najskuteczniej, gdy zostaje włączona w codzienne praktyki samych podmiotów. Challenger uzupełnia tę tradycję o współczesny aparat biologiczny: autonomię można osłabić poprzez kształtowanie środowiska bodźców, w którym nasze systemy regulacyjne, społeczne i afektywne zaczynają generować zachowania przewidywalne dla zewnętrznego projektanta. W takim świecie człowiek nie jest zmuszany do kliknięcia – zostaje otoczony środowiskiem, w którym coraz częściej chce to zrobić. I tutaj dopiero zaczyna się prawdziwy dramat agency hacking.

Od panoptikonu do psychografii bodźca. Panoptikon Benthama, rozsławiony przez Foucaulta, opierał się na możliwości obserwacji; więzień, nie wiedząc, czy jest w danej chwili oglądany, uczył się zachowywać tak, jakby nadzór trwał stale. Cyfrowe środowisko wykonuje ruch znacznie subtelniejszy. Ono nie tylko obserwuje – zbiera dane, by coraz trafniej przewidywać, który bodziec zmieni przyszłe zachowanie. Różnica jest zasadnicza.

Klasyczny obserwator pyta: „co robisz?". System predykcyjny pyta dodatkowo: „co najprawdopodobniej zrobisz, jeśli pokażę ci A zamiast B?". Gdy ta zdolność przewidywania łączy się z możliwością natychmiastowej modyfikacji środowiska informacyjnego, obserwacja zmienia się w pętlę sterowania. Człowiek reaguje na bodziec, reakcja staje się daną, a dana pozwala dobrać kolejny bodziec, co jeszcze bardziej personalizuje otoczenie. To już nie jest zwykła informacja o użytkowniku – to eksperyment prowadzony w czasie rzeczywistym na jego uwadze. Challengerowskie crazescapes należy rozumieć właśnie jako antropogeniczny odpowiednik wcześniejszych crave-scapes.

Podobnie jak przemysł żywności ultraprzetworzonej nauczył się wykorzystywać fizjologiczne mechanizmy apetytu, tak gospodarka cyfrowa uczy się eksploatować mechanizmy uwagi, nagrody, ciekawości, przynależności i oburzenia. Oba te środowiska są przykładami systemów zaprojektowanych tak, by rozregulowywać biologiczne i poznawcze mechanizmy samokontroli. W obu przypadkach problemem nie jest to, że człowiek traci wolę, lecz to, że środowisko zostaje zoptymalizowane pod przewidywalne słabości tej woli.

Neurorights jako ostatnia granica ochrony ludzkiej autonomii

Neurorights: prawa człowieka docierają do wnętrza umysłu. Jeśli historia praw człowieka była procesem stopniowego poszerzania obszaru chronionego przed arbitralną ingerencją, rozwój neurotechnologii stawia pytanie o nową granicę: integralność psychiczną, mentalną prywatność, ciągłość osobowej tożsamości i wolność myśli. Materiał Challenger przywołuje właśnie kategorię neurorights, łącząc ją z potrzebą ochrony psychosfery przed technologiami zdolnymi wpływać na wewnętrzne warunki sprawstwa. To już nie jest wyłącznie futurystyczne słownictwo. W listopadzie 2025 roku UNESCO przyjęło Recommendation on the Ethics of Neurotechnology — pierwszy globalny standard normatywny tej organizacji dotyczący etyki neurotechnologii, obejmujący między innymi kwestie godności, autonomii, tożsamości i prywatności mentalnej.

Humanistyczna doniosłość tego ruchu jest ogromna. Liberalne prawo przez stulecia budowało ochronę domu, ciała, korespondencji, sumienia i wypowiedzi. Neurotechnologia zmusza do pytania, czy ostatnim "domem" człowieka jest jego własny umysł. Jeżeli można coraz dokładniej wnioskować o stanach mentalnych lub wpływać na procesy neuronalne, klasyczne pojęcie prywatności staje się niewystarczające. UNESCO już w pracach przygotowawczych wskazywało właśnie na integralność mentalną, autonomię, ciągłość psychologiczną i prywatność umysłu jako szczególne obszary ryzyka. W tym miejscu Challenger spotyka Johna Stuarta Milla.

Liberalizm klasyczny chronił sferę, w której człowiek może kształtować własny sposób życia, dopóki nie narusza innych. XXI wiek dodaje pytanie o stopień wcześniejszy: kto kształtuje mechanizmy, z których później rodzi się "własny sposób życia"? Jeśli możliwe staje się coraz precyzyjniejsze modelowanie oraz oddziaływanie na uwagę, nastrój, pamięć czy decyzję, wolność nie może być definiowana jedynie jako brak fizycznego przymusu. Można pozostawić człowieka formalnie wolnym, a jednocześnie coraz doskonalej projektować warunki jego wyboru. To problem, którego Monteskiusz, Locke i Mill nie mogli przewidzieć w tej technologicznej postaci, ale którego ich podstawowe pytanie dotyczyło od początku: gdzie kończy się prawomocna władza nad człowiekiem?

Polemika z Challenger: organizm nie ma monopolu na przyszłość. Po tej rozbudowanej humanistycznej rekonstrukcji trzeba jednak ponownie przeciwstawić się Challenger w punkcie dla niej najważniejszym. Jej krytyka współczesnej AI jest przekonująca, gdy twierdzi, że zdolność do generowania ludzkopodobnych zachowań nie dowodzi istnienia świadomości, że obecne systemy nie wykazują biologicznej autonomii organizmu oraz że język "inteligencji" może skłaniać do pochopnego antropomorfizowania artefaktów. Współczesna literatura naukowa rzeczywiście nie daje podstaw do prostego stwierdzenia, że obecnie używane systemy AI są świadome; badacze proponują ostrożne metody oceny zamiast wnioskowania z samej elokwencji systemu.

Znacznie słabsza staje się jednak jej teza wtedy, gdy "brak metabolizmu biologicznego" zaczyna pełnić funkcję metafizycznego dowodu niemożliwości wszelkiej sztucznej podmiotowości. Nie wiemy, czy świadomość wymaga dokładnie biologii komórkowej, czy też pewien typ zorganizowanej, ucieleśnionej, samopodtrzymującej i globalnie zintegrowanej dynamiki mógłby zostać zrealizowany na innym substracie. Współczesny spór naukowy i filozoficzny nie jest w tej kwestii zamknięty; właśnie dlatego autorzy zajmujący się badaniami nad świadomością AI postulują ostrożność również w drugą stronę — aby nie stworzyć potencjalnie świadomych systemów bez refleksji nad ich statusem moralnym.

Brak świadomości maszyny nie wyklucza jej niszczycielskiej skuteczności

Dziedzictwem Challengera nie powinien być zatem dogmat, że maszyna nigdy nie będzie podmiotem, lecz ostrzejsze kryterium samej podmiotowości. Nie pytajmy wyłącznie o to, czy system mówi inteligentnie. Pytajmy, czy posiada własną perspektywę i historię stanów mających dla niego znaczenie; czy dysponuje autonomicznymi mechanizmami utrzymywania organizacji oraz zdolnością ustanawiania celów, których nie da się wyczerpać w intencji projektanta. I wreszcie – co najtrudniejsze – czy istnieje cokolwiek, co z jego własnego punktu widzenia jest dobre lub złe. To test wymagający zarówno dla maszyny, jak i dla naszej filozofii.

Najgroźniejsza AI nie musi być świadoma. Tu dochodzimy do paradoksu, który politycznie może być ważniejszy niż samo pytanie o świadomość maszyn. System nie potrzebuje świadomości, by wpływać na świadomość człowieka. Nie musi pragnąć władzy, aby stać się jej narzędziem. Nie musi rozumieć słabości użytkownika, jeśli wystarczy, że statystycznie przewiduje bodźce skorelowane z określoną reakcją. Nie potrzebuje własnego telos, by niezwykle skutecznie realizować cel instytucji, która go wdraża. W tym miejscu termin "Artificial Obedience" okazuje się bardziej przenikliwy, niż sugerowałaby jego retoryczna przesada. Posłuszna maszyna może być społecznie potężniejsza od autonomicznej, o ile jej ogromna kompetencja zostanie podporządkowana zewnętrznemu celowi: zyskowi, wpływowi, kontroli, propagandzie, selekcji lub optymalizacji zaangażowania.

Brak własnej moralnej intencji nie stanowi zabezpieczenia. Gilotyna również nie nienawidzi człowieka. Humanistyka zna ten problem doskonale; najgroźniejsze systemy w historii rzadko działały wyłącznie dzięki namiętności jednej osoby. Ich siła wynikała z aparatu procedur, biurokracji, technologii, norm i rozproszenia odpowiedzialności. Max Weber dostrzegał "żelazną klatkę" racjonalizacji, Kafka stworzył obraz człowieka uwięzionego w procedurze, a Hannah Arendt badała zło wykonywane w rytmie obowiązku administracyjnego. AI wprowadza do tej historii nową możliwość: procedura może teraz odpowiadać, przewidywać i personalizować. Biurokracja otrzymuje język. Rynek otrzymuje predykcję. Propaganda otrzymuje indywidualnego rozmówcę. To nadal nie dowodzi świadomości maszyny, ale pokazuje, dlaczego wcale nie musi być ona pierwszym problemem, o który powinniśmy pytać.

Człowiek jako ostatnie pole eksploatacji. Od Marksa po szkołę frankfurcką humanistyka uczyła, że kapitalizm stale poszukuje nowych obszarów do włączenia w proces wartościowania ekonomicznego. Eksploatowane mogą być ziemia, czas pracy, surowce, relacje społeczne czy kultura. W gospodarce cyfrowej szczególnie cennym zasobem staje się przewidywalność zachowań. Jeśli można coraz trafniej przewidzieć, na co człowiek zwróci uwagę, co kupi, komu zaufa, czego się

przestraszy i kiedy straci zainteresowanie, jego podmiotowość staje się polem ekonomicznej ekstrakcji. Challengerowskie agency hacks można zatem połączyć z krytyką kapitalizmu uwagi: system nie sprzedaje już wyłącznie przedmiotu konsumentowi; coraz częściej optymalizuje środowisko, w którym ten konsument zaczyna czegoś chcieć. To istotna zmiana perspektywy. Stara reklama mówiła: "kup mój produkt".

Hakowanie sprawstwa jako manipulacja warunkami wyłaniania się ludzkich celów

Nowa infrastruktura może najpierw badać, jaki układ bodźców sprawi, że wezwanie stanie się skuteczniejsze właśnie wobec ciebie. Wtedy pragnienie staje się jednocześnie doświadczeniem pierwszoosobowym i zmienną modelu. To być może najbardziej charakterystyczna dla XXI wieku forma alienacji: człowiek nadal czuje „ja chcę", ale ktoś inny potrafi coraz dokładniej modelować warunki, w których owo „chcę" się pojawi. Chodzi o ocalenie „dlaczego". W tym miejscu cały łuk Alive zaczyna ukazywać swoją humanistyczną logikę. Nasiono sekwoi, komórka, rekin polarny, ośmiornica, śluzowiec, pstrąg i człowiek byli dla Challenger wariantami jednego pytania: gdzie znajduje się własne „dlaczego" żyjącej istoty? Teraz pytanie to staje się polityczne.

Jeśli człowiek posiada własne cele, ale techniczne środowisko potrafi coraz skuteczniej je modulować, ochrona podmiotowości nie może ograniczać się do ochrony życia biologicznego. Musi obejmować warunki, w których cele mogą rzeczywiście być w dostatecznym stopniu nasze. Arystoteles nazwałby to problemem dobrego życia, Kant autonomią, Mill indywidualnością, Arendt wolnością działania, Foucault relacją podmiotu i władzy, Stiegler troską o techniczne warunki indywiduacji, a Challenger – ochroną ucieleśnionego sprawstwa. Różnią ich epoki, języki i założenia, lecz wszyscy w pewnym sensie pytają o to samo: co sprawia, że człowiek nie jest jedynie miejscem, przez które przepływają siły silniejsze od niego? Nie wystarcza odpowiedzieć: możliwość wyboru. Trzeba jeszcze zapytać, kto zbudował menu. Czas zejść głębiej w politykę psychiki: w agency hacking, pseudo-wspólnoty, epidemiologię oburzenia, ekonomię uwagi, wolność kognitywną oraz neurorights. Tam trzeba będzie powrócić do Nicholasa Humphreya i społecznej genezy samoświadomości, ponieważ najbardziej podatnym miejscem człowieka okazuje się nie jego samotność, lecz potrzeba przynależności. Challenger właśnie z tej społecznej otwartości wyprowadza szczególną podatność na fake-groups i crazescapes. Człowiek jest zwierzęciem, które nauczyło się widzieć siebie w oczach drugiego człowieka. Dlatego można go także skłonić do patrzenia w zwierciadło, za którym nie stoi wspólnota, lecz mechanizm optymalizacji.

W tej chwili pytanie Alive przestaje brzmieć wyłącznie: czym jest życie? Coraz wyraźniej brzmi: kto będzie miał prawo programować środowisko, w którym życie wybiera swoje cele? Hakowanie sprawstwa: od neurobiologii nagrody do politycznej architektury uwagi. Doszliśmy do paradoksu, który stanowi jedno z najbardziej niepokojących odkryć współczesnej kultury cyfrowej: maszyna nie musi posiadać własnego telos, aby skutecznie uczestniczyć w kształtowaniu telos człowieka. Nie musi odczuwać pragnienia, aby przewidywać, co będzie dla nas pożądane; nie musi doświadczać gniewu, aby rozpoznawać treści zdolne go wywołać; nie musi posiadać potrzeby przynależności, aby klasyfikować sygnały wzmacniające identyfikację grupową. Melanie Challenger określa tę podatność mianem agency hacking, „hakowania sprawstwa", i wiąże ją z fake-groups, dezinformacją, crazescapes oraz coraz bardziej wyrafinowanymi środowiskami technologicznymi zdolnymi ingerować w warunki, z których wyłania się ludzkie działanie. Warto to pojęcie od razu oczyścić z nadmiernej retoryki.

Techniczne hakowanie biologicznych pętli nagrody

"Hakowanie" nie oznacza, że algorytm przejmuje kontrolę nad człowiekiem w sposób, w jaki napastnik przejmuje komputer. W większości przypadków mamy do czynienia z mechanizmem znacznie subtelniejszym: system techniczny może zmieniać rozkład prawdopodobieństwa naszych zachowań poprzez modyfikację kolejności, częstotliwości, emocjonalną temperaturę i społeczną widoczność bodźców. Z perspektywy nauk behawioralnych jest to kwestia uczenia ze wzmocnieniem, nawyku, saliencji, predykcji nagrody, społecznego uczenia się oraz regulacji uwagi; z punktu widzenia filozofii politycznej pojawia się pytanie, czy wolność można zachować formalnie, a jednocześnie systematycznie przekształcać środowisko, w którym ten wolny wybór powstaje. Nagroda społeczna nie jest metaforą. Biologiczne podstawy tego problemu są stosunkowo dobrze rozpoznane.

Ludzkie zachowanie podlega uczeniu na podstawie konsekwencji, a nagroda społeczna pełni funkcję wzmacniającą podobnie jak inne bodźce. Analiza ponad miliona postów od czterech tysięcy użytkowników wykazała, że częstotliwość publikowania w mediach społecznościowych można opisać za pomocą modeli reinforcement learning: wcześniejsze nagrody społeczne wpływały na późniejszą aktywność, co potwierdził eksperyment internetowy wykazujący przyczynowy wpływ społecznego wzmocnienia na zachowanie użytkowników. Nie dowodzi to, że platforma jest zwykłym "pudełkiem Skinnera" ani że każdy z nas działa według identycznego algorytmu, ale potwierdza kluczową dla Challengera tezę: cyfrowe środowisko może wejść w zamkniętą pętlę z biologicznymi mechanizmami uczenia się człowieka. Pętla ta wydaje się prosta tylko na pierwszy rzut oka.

Człowiek wykonuje działanie, otrzymuje odpowiedź społeczną, aktualizuje oczekiwanie dotyczące przyszłej nagrody, a następnie zmienia prawdopodobieństwo powtórzenia danego zachowania. W języku uczenia ze wzmocnieniem kluczową rolę odgrywa reward prediction error – różnica między nagrodą oczekiwaną a rzeczywistą, która służy aktualizacji wartości przypisywanej działaniom i stanom. Oczywiście w realnym życiu mechanizmy te współistnieją z pamięcią, celami, normami społecznymi, nawykami i refleksyjną kontrolą, dlatego człowieka nie da się zredukować do prostego algorytmu reinforcement learning; współczesne modele badają właśnie interakcję uczenia nagrodowego, pamięci oraz mechanizmów nawykowych. Znaczenie tego procesu dla mediów społecznościowych polega na stworzeniu układu, w którym reakcje innych ludzi stają się jednocześnie doświadczeniem społecznym i mierzalnym sygnałem technicznym. Polubienie, udostępnienie, liczba obserwujących czy wzrost widoczności są przeżywane jako informacja o uznaniu, statusie lub przynależności, a zarazem stają się danymi dla systemu rekomendacyjnego. Żywe środowisko społeczne zostaje częściowo przetłumaczone na metryki, które z kolei zaczynają zwrotnie kształtować zachowania społeczne. W języku Challengera można powiedzieć, że technika nie tylko obserwuje psyche; wchodzi w jego obieg regulacyjny.

Cyfrowa architektura nawyków i mechanizmy selekcji emocjonalnej

Arystoteles spotyka uczenie ze wzmocnieniem. W tym miejscu niezwykle użyteczna okazuje się etyka Arystotelesa. Dla autora „Etyki nikomachejskiej” charakter nie był prywatną substancją ukrytą w człowieku, lecz rezultatem praktyki: stajemy się określonymi ludźmi poprzez powtarzane działania, a dyspozycje utrwalają się w nawykach. Współczesna psychologia uczenia się posługuje się zupełnie innym aparatem pojęciowym, lecz w jednym punkcie starożytna intuicja i nowoczesna nauka niepokojąco się spotykają: to, co często robimy, zmienia prawdopodobieństwo tego, co będziemy robić później.

Cyfrowa architektura zachowania zmienia zatem znaczenie dawnego pytania o kształtowanie charakteru. Kiedyś środowisko nawyku tworzyły przede wszystkim rodzina, szkoła, wspólnota, obyczaj, rytuał i bezpośrednie konsekwencje działania. Dzisiaj dołączyły do nich systemy zdolne na bieżąco obserwować zachowania milionów użytkowników oraz reorganizować środowisko bodźców. Nie oznacza to, że algorytm „wychowuje” w takim samym sensie jak nauczyciel, ale wskazuje, że powtarzalność ludzkich reakcji może być współkształtowana przez architekturę, która nie jest neutralnym kanałem przekazu. Jeżeli Arystoteles uważał, że dobre państwo powinno troszczyć się o

warunki kształtowania cnót obywateli, XXI wiek musi zadać pytanie wcześniejsze: kto dzisiaj projektuje warunki kształtowania nawyków?

Problem nie zaczyna się od treści pojedynczej opinii, lecz od rytmu sprawdzania powiadomień, oczekiwania na reakcję, przenoszenia uwagi, wartościowania własnej widoczności i uczenia się tego, jakie formy ekspresji przynoszą największy odzew społeczny. Emocja może stać się parametrem dyfuzji. W źródle Challenger pojawia się odwołanie do eksperymentu Facebooka nad „zaraźliwością emocjonalną". Warto sprostować jego chronologię i zakres: szeroko dyskutowane badanie opublikowano w 2014 roku, a objęło ono 689 003 użytkowników. Manipulując proporcją pozytywnych i negatywnych wpisów w strumieniu treści, badacze wykazali niewielkie, lecz statystycznie mierzalne przesunięcia emocjonalnego tonu późniejszych postów użytkowników.

Eksperyment stał się ważnym precedensem metodologicznym i etycznym, pokazując, że infrastruktura komunikacyjna może służyć jako środowisko oddziaływania na skalę nieosiągalną dla klasycznego laboratorium psychologicznego. Jeszcze istotniejsze dla koncepcji crazescapes są badania wykazujące, że określone właściwości emocjonalne komunikatu korelują z jego transmisją. Analiza języka moralno-emocjonalnego w dyskusjach politycznych wykazała, że obecność słów łączących treść moralną z emocją zwiększa prawdopodobieństwo rozpowszechniania komunikatów wewnątrz sieci ideologicznych. Późniejsze badania nad wrogością wobec grup przeciwnych wskazywały, że język odnoszący się do politycznego out-group może szczególnie silnie wiązać się z zaangażowaniem; badanie przeprowadzone w dziewięciu krajach, w tym w Polsce, wykazało systematycznie większą toksyczność interakcji między przeciwnymi obozami niż kontaktów wewnątrzgrupowych.

Szczególnie wymowne są wyniki dotyczące oburzenia. Badanie opublikowane w Science w 2024 roku wykazało, że dezinformacja częściej wywołuje oburzenie, a to z kolei zwiększa prawdopodobieństwo jej udostępniania; autorzy analizowali ten związek w wielu zbiorach danych i przeprowadzili eksperymenty dotyczące intencji dzielenia się treściami. Nie należy z tego wyprowadzać naiwnej tezy, że „algorytmy powodują gniew", ponieważ oburzenie jest realnym ludzkim afektem występującym niezależnie od platform cyfrowych, a użytkownicy sami wybierają i wytwarzają treści. Wynik ten obnaża jednak mechanizm selekcyjny: jeżeli środowisko nagradza zasięg i zaangażowanie, treść posiadająca właściwości sprzyjające transmisji może otrzymywać przewagę reprodukcyjną w ekosystemie informacji.

Tutaj humanistyka może przywołać starą obserwację retoryki Arystotelesa: perswazja nigdy nie była wyłącznie operacją na argumentach, gdyż logos współdziałał z ethos i pathos. Nowość technologiczna polega nie na odkryciu emocji w polityce, lecz na możliwości kwantyfikacji reakcji emocjonalnych w czasie niemal rzeczywistym i używania ich jako zmiennych optymalizacji. To, co

dawny mówca intuicyjnie wyczuwał z reakcji tłumu, współczesna infrastruktura mierzy poprzez kliknięcia, komentarze, oglądalność, retencję i dalszą dyfuzję.

Nie demonizujmy algorytmu: użytkownik również dokonuje wyborów. Profesjonalna analiza wymaga jednak odrzucenia determinizmu technologicznego. Gdyby algorytm samodzielnie „produkował polaryzację", stosunkowo prosta zmiana feedu powinna skutecznie zmieniać przekonania. Duży eksperyment na Facebooku przeprowadzony podczas wyborów w USA w 2020 roku pokazał coś bardziej złożonego: obniżenie ekspozycji na treści ze źródeł zgodnych politycznie z użytkownikiem o około jedną trzecią ograniczyło kontakt z częścią treści nieobyczajnych i źródłami wielokrotnie rozpowszechniającymi dezinformację, lecz nie przyniosło mierzalnej zmiany w ośmiu zarejestrowanych wskaźnikach postaw politycznych, w tym polaryzacji afektywnej i ekstremizmu ideologicznego. To niezwykle ważna korekta wobec zbyt efektownej wersji agency hacking.

Człowiek nie jest plasteliną, którą algorytm może dowolnie modelować; społeczne przekonania mają wiele źródeł: rodzinę, doświadczenie, media tradycyjne, środowisko zawodowe, religię, klasę społeczną, historię polityczną i istniejące tożsamości. Algorytmiczne środowisko wchodzi w istniejący system, a nie tworzy człowieka od zera. W języku biologii Challenger oznacza to, że psyche zachowuje własną historię i strukturę, a bodziec zawsze trafia do organizmu już ukształtowanego przez wcześniejsze doświadczenia.

Algorytmy rządzą poprzez konfigurację pola uwagi

Nie wolno jednak na podstawie tych wyników wyciągać równie uproszczonego wniosku, jakoby algorytmy były politycznie obojętne. Randomizowany eksperyment dotyczący feedu platformy X, opublikowany w Nature w 2026 roku, wykazał, że siedmiotygodniowe korzystanie z rekomendacji algorytmicznych przesuwało część postaw użytkowników z USA w kierunku bardziej konserwatywnym. Jednym z obserwowanych mechanizmów było zwiększenie ekspozycji na określone treści i późniejsze śledzenie odpowiednich kont, choć nie stwierdzono przy tym znaczącej zmiany identyfikacji partyjnej ani ogólnego poziomu polaryzacji. Nauka zmusza nas zatem do zastąpienia sloganu „algorytmy sterują ludźmi" sformułowaniem znacznie bardziej precyzyjnym: algorytmiczne uporządkowanie środowiska informacyjnego może w określonych warunkach przyczynowo zmieniać ekspozycję, zachowania i niektóre postawy, jednak skala oraz kierunek tego efektu zależą od platformy, konstrukcji systemu, konkretnego użytkownika, czasu ekspozycji i badanego rezultatu. Teza ta jest mniej widowiskowa niż teoria cyfrowego zniewolenia, lecz właśnie dlatego naukowo bardziej niebezpieczna: wskazuje na realny, mierzalny wpływ bez potrzeby

kreowania mitu o wszechmocnym algorytmie. W tym punkcie Melanie Challenger spotyka się z Michelem Foucaultem i jego koncepcją władzy jako projektowania pola możliwego działania.

W późniejszych analizach Foucaulta władza nie sprowadza się do prostych zakazów i nakazów; governmentality pozwala badać techniki organizowania zachowań populacji poprzez określone racjonalności, statystyki, instytucje i warunki działania. Władza może operować nie przez odbieranie możliwości, lecz przez konfigurowanie pola, w którym pewne zachowania stają się łatwiejsze, bardziej naturalne lub prawdopodobne od innych. Algorytm rekomendacyjny jest z tej perspektywy interesujący nie jako nowy suweren, lecz jako nadzwyczaj precyzyjne narzędzie rządzenia widzialnością. Nie musi on zakazywać lektury artykułu, którego użytkownik po prostu nie widzi; wystarczy, że inny tekst otrzyma pierwszeństwo. Nie musi nakazywać wejścia w dyskusję – może jedynie zwiększyć dostępność komunikatu, który historycznie częściej wywoływał reakcję. Nie musi usuwać alternatyw, by zmienić ich względną obecność w polu uwagi.

To właśnie odróżnia miękką władzę infrastruktury od klasycznej cenzury. Cenzor mówi: „tego nie wolno zobaczyć". System selekcji szepcze niemal niesłyszalnie: „to zobaczysz najpierw". Dla organizmu biologicznego, ograniczonego czasem, pamięcią roboczą i uwagą, różnica w kolejności prezentacji treści nie jest niewinna. Jak zauważył Herbert Simon: bogactwo informacji rodzi ubóstwo uwagi. W naukach o decyzji Simon rozwinął koncepcję ograniczonej racjonalności (bounded rationality): człowiek, nie dysponując nieograniczonym czasem, pełną informacją ani nieskończoną mocą obliczeniową, podejmuje decyzje za pomocą heurystyk i rozwiązań „wystarczająco dobrych", zamiast dążyć do wszechwiedzącej maksymalizacji. Ma to kluczowe znaczenie w środowisku informacyjnym, gdzie nadmiar bodźców sprawia, że problemem przestaje być dostęp do informacji, a staje się nim alokacja ograniczonej uwagi.

Filozoficzne i ekonomiczne analizy racjonalności wciąż traktują Simonowską ograniczoną racjonalność jako jeden z głównych sposobów odejścia od modelu idealnie racjonalnego decydenta. Z perspektywy Challenger jest to wręcz fizjologiczne przedłużenie teorii homeostazy. Uwaga nie jest nieskończonym światłem, którym świadomość mogłaby oświetlić wszystko jednocześnie; organizm musi selekcjonować, gdyż jego zasoby poznawcze są skończone. Każde środowisko informacyjne konkuruje zatem nie tylko o przekonania użytkownika, lecz o czas biologicznego systemu, który może zostać poświęcony jednemu bodźcu kosztem innego. Ekonomia uwagi jest więc ekonomią w najbardziej dosłownym sensie: dotyczy dobra rzadkiego. Minuta spędzona na jednym komunikacie nie może być w tej samej chwili przeznaczona na rozmowę, sen, pracę, książkę, spacer czy inny komunikat. Jeśli rynek nauczy się skuteczniej niż jednostka przewidywać, co zatrzyma jej uwagę, powstaje asymetria wiedzy. Organizm posiada uwagę; system coraz lepiej zna warunki jej przechwytywania. Nawyk nie jest tożsamy z uzależnieniem.

Hakowanie sprawstwa poprzez nawyki i potrzebę uznania

W tym miejscu konieczna jest korekta naukowa wobec popularnej retoryki. Nie każde intensywne korzystanie z mediów społecznościowych jest uzależnieniem w znaczeniu klinicznym, a sam język "dopaminowego uzależnienia od telefonu" bywa znacznie bardziej pewny siebie niż literatura przedmiotu. Badania z 2025 roku wskazują, że część użytkowników może przeceniać stopień swojego "uzależnienia", podczas gdy ich zachowanie lepiej opisują utrwalone nawyki; autorzy zwracają uwagę, że pochopne nadawanie etykiety uzależnienia może paradoksalnie zmniejszać poczucie kontroli u samych użytkowników. Dla argumentu Challenger ta korekta jest kluczowa.

Jeżeli wszystko nazwiemy uzależnieniem, stracimy możliwość rozpoznania subtelniejszych form heteronomii. Człowiek nie musi cierpieć na kliniczne zaburzenie, aby jego dzień został pocięty na setki małych reakcji na bodźce; nie musi tracić całkowitej kontroli, by nawyk sprawdzania urządzenia zawężał przestrzeń dla refleksyjnego wyboru. Hakowanie sprawstwa może zatem polegać nie na spektakularnym zniewoleniu, lecz na mikroskopijnej kolonizacji sekwencji dnia. Właśnie dlatego nauki o nawykach są tu być może ważniejsze od języka patologii. Zachowanie wielokrotnie uruchamiane w podobnym kontekście z czasem staje się coraz mniej zależne od świadomej deliberacji. Technologia nie musi więc "przejąć mózgu"; wystarczy, że stanie się niezawodnym elementem kontekstu, w którym konkretna reakcja została wielokrotnie wzmocniona.

Tu pojawia się kwestia Nicholasa Humphreya i biologicznej ceny społeczności. Źródło Challenger łączy podatność na cyfrowe manipulacje z argumentem Humphreya o społecznych korzeniach rozwiniętej samoświadomości. Jeśli ludzki umysł został ukształtowany przez konieczność życia w złożonych grupach, to sygnały statusu, reputacji, akceptacji czy odrzucenia nie są powierzchownymi dodatkami do "racjonalnego człowieka". Stanowią one integralną część środowiska, do którego nasza psychologia była dostrajana ewolucyjnie; notatka Alive wiąże zresztą społeczną otwartość człowieka z podatnością na pseudo-grupy i algorytmiczne hakowanie sprawstwa. Badania nad rozwojem mózgu wskazują, że szczególnie w okresie adolescencji kluczowe znaczenie nabierają akceptacja rówieśnicza i nagrody społeczne, a układy związane z nagrodą reagują na sygnały popularności, takie jak "lajki". Jednocześnie autorzy przeglądów podkreślają ogromne różnice indywidualne i przestrzegają przed wyprowadzaniem prostych relacji przyczynowych między czasem ekranowym a rozwojem mózgu. W perspektywie humanistycznej oznacza to jednak coś więcej niż tylko podatność młodzieży na ekrany.

Człowiek potrzebuje uznania przez innych, ponieważ jego tożsamość jest relacyjna. Hegel budował na tym dialektykę uznania, George Herbert Mead opisywał "ja" poprzez internalizację relacji społecznych, a współczesna psychologia społeczna dowodzi, że normy i oczekiwania grupowe realnie kształtują zachowanie. Filozofia norm podkreśla również rolę wzajemnych oczekiwań i wspólnej wiedzy w utrzymywaniu reguł społecznych.

Przemysłowa fabrykacja istotności i erozja świata wspólnego

Problem fake-groups nie polega na tym, że wspólnoty internetowe są „nierzeczywiste". Przyjaźń, solidarność czy realna organizacja społeczna mogą przecież zachodzić w przestrzeni cyfrowej. Problem zaczyna się wtedy, gdy sygnały wspólnotowości zostają technicznie wykorzystane bez odpowiadającej im głębi relacji, a jednostka otrzymuje silne bodźce identyfikacyjne od systemu, którego struktur i selekcji nie rozumie. Wróg jako wyjątkowo skuteczny klej społeczny.

Badania nad politycznym zaangażowaniem online odsłaniają szczególnie niebezpieczną wersję tego procesu. Analizy treści wskazują, że odwołania do wrogiej grupy zewnętrznej bywają silnym predyktorem udostępnień; tożsamość polityczna może być komunikowana nie tylko poprzez deklarowanie miłości do własnego obozu, lecz właśnie przez niechęć wobec przeciwnika. Nie oznacza to, że media społecznościowe wynalazły trybalizm – Tukidydes, Machiavelli czy Hobbes oraz cała historia konfliktów politycznych świadczą o czymś przeciwnym.

Nowością jest jednak możliwość przemysłowego mierzenia reakcji na komunikaty i ich natychmiastowej dystrybucji w ogromnych sieciach. Dawny demagog musiał intuicyjnie oceniać tłum; współczesny system operuje strumieniem danych o tym, co zostało zatrzymane, udostępnione lub polubione. Challengerowski crazescape można zatem zdefiniować naukowo ostrożniej niż czyni to retoryka źródła: nie jako świat, w którym algorytm „doprowadza ludzi do szaleństwa", lecz jako środowisko informacyjne, w którym struktura bodźców, mechanizmy społecznego uczenia i funkcje optymalizacji mogą systematycznie premiować określone rodzaje treści. Jeżeli oburzenie, tożsamościowa antagonizacja lub sensacyjność generują większe zaangażowanie, system nastawiony na engagement może wytworzyć presję selekcyjną sprzyjającą właśnie takim komunikatom. Związek ten wymaga badania platforma po platformie i nie może być przyjmowany jako uniwersalne prawo, lecz istnieją już empiryczne dane, by traktować problem poważnie. Arendt przypomina: społeczeństwo potrzebuje świata wspólnego, a nie tylko wielu feedów.

W tym miejscu "hakowanie sprawstwa" staje się problemem politycznym w sensie Hannah Arendt. Wiązała ona działanie polityczne z pluralnością i wspólnym światem, w którym ludzie mogą ukazywać się sobie nawzajem poprzez słowa i czyny; jej diagnoza nowoczesności obejmowała lęk przed utratą publicznego świata na rzecz konformizmu, administracji i manipulacji opinią. Algorytmiczna personalizacja nie musi niszczyć takiego świata, ale może zmienić warunki jego powstawania. Jeśli dwaj obywatele otrzymują radykalnie odmienne uporządkowanie informacji na podstawie historii zachowań, problemem nie jest samo zróżnicowanie perspektyw – pluralność była dla Arendt warunkiem polityki. Problem pojawia się, gdy znika wystarczająca liczba wspólnych przedmiotów odniesienia. Wtedy spór przestaje być sporem o ten sam świat, a zaczyna przypominać równoległe monologi prowadzone na odmiennych zbiorach faktów i hierarchiach ważności. Tutaj dane naukowe wymagają powściągliwości.

Słynna hipoteza prostych „baniek filtrujących" nie została potwierdzona w swojej najbardziej katastroficznej wersji: duże eksperymenty na Facebooku wykazały znaczną ekspozycję na źródła podobne ideologicznie, lecz samo zmniejszenie tej ekspozycji nie przyniosło automatycznej redukcji polaryzacji. Oznacza to, że erozja wspólnego świata nie wynika z jednego mechanizmu rekomendacji. Jest to zjawisko społeczne, polityczne i kulturowe, w którym platformy są jednym z uczestników, a nie demiurgiem. Crazescape jako środowisko allostatyczne pozwala powiązać kulturę cyfrową z fizjologią opisaną we wcześniejszych częściach artykułu.

Allostaza to stabilność osiągana poprzez zmianę i zdolność organizmu do dostosowywania reakcji do przewidywanego zapotrzebowania. Jeśli jednak środowisko bez przerwy dostarcza sygnałów nowości, zagrożenia czy konieczności reakcji, układ regulacyjny może być stale angażowany przez wydarzenia, z których każde osobno wydaje się małe, lecz których łączna częstotliwość tworzy nową sytuację. Nie należy z tego tworzyć prostego równania „telefon = stres". Zależności między mediami cyfrowymi a zdrowiem psychicznym są heterogeniczne i zależą od sposobu używania technologii; literatura rozróżnia wzorce patologiczne od częstego, lecz kontrolowanego użycia.

Challenger może jednak postawić pytanie fizjologiczne, którego nie trzeba rozstrzygać moralną paniką: co dzieje się z organizmem, którego krajobraz relewancji jest nieustannie przerywany przez konkurujące sygnały? W naturalnym środowisku saliencja była cenna, bo wskazywała coś istotnego: zagrożenie, pokarm, partnera czy płacz dziecka. W środowisku technicznym saliencja może być produkowana celowo, ponieważ to, co wyróżniające się, ma większą szansę zdobyć uwagę. Ewolucyjnie stary mechanizm "zwróć uwagę na to, co istotne" spotyka przemysł zdolny do fabrykowania "istotności".

W tej perspektywie crazescape jest kuzynem crave-scape. Pierwszy eksploatuje architekturę uwagi i społecznego afektu, drugi architekturę apetytu i nagrody metabolicznej. Oba działają skutecznie nie

dlatego, że organizm jest źle zaprojektowany, lecz dlatego, że jego adaptacje zostają przeniesione do środowiska o właściwościach nieobecnych w warunkach, w których kształtowały się jego mechanizmy regulacyjne.

Dezinformacja i agency hacking jako techniczne sterowanie biologicznym zaufaniem

Dezinformacja jako pasożytnictwo na epistemologii społecznej. Problem fałszywej informacji staje się znacznie ciekawszy, gdy porzucimy moralizatorskie pytanie: „dlaczego ludzie wierzą w głupoty?". Wiedza człowieka od zawsze miała charakter społeczny. Nikt z nas nie weryfikuje samodzielnie ruchu planet, struktury atomu, bezpieczeństwa każdego mostu czy składu każdego leku.

Epistemologia wspólnotowa opiera się na zaufaniu, reputacji, autorytecie, świadectwie i podziale pracy poznawczej. Bez tych mechanizmów cywilizacja naukowa byłaby niemożliwa. Dezinformację można zatem rozumieć jako pasożyta wykorzystującego mechanizmy niezbędne do zdobywania wiedzy. Człowiek ufa innym, ponieważ nie jest w stanie sprawdzić wszystkiego osobiście; manipulator instrumentalnie wykorzystuje właśnie tę ekonomię zaufania. Zwracamy uwagę na ostrzeżenia, gdyż informacje o zagrożeniu mają wysoką wartość biologiczną – fałszywy komunikat może przejąć ten mechanizm. Szczególną wagę przypisujemy świadectwom własnej grupy; propaganda potrafi zatem przybrać maskę wspólnotowej lojalności. Badania nad przepływem informacji wskazują, że emocjonalne oburzenie osłabia związek między skłonnością do udostępnienia treści a jej dokładnością, co staje się paliwem dla fałszywych przekazów.

Nie jest to zatem wojna „rozumu z emocjami", gdyż emocje stanowią integralną część ludzkiej racjonalności praktycznej. To raczej konflikt między mechanizmami poznawczymi, dostosowanymi do decyzji w świecie ograniczonej informacji, a systemami zdolnymi masowo testować, które komunikaty te mechanizmy najskuteczniej uruchamiają.

Od manipulacji do cybernetyki człowieka. W tym miejscu powraca postać Norberta Wienera, choć teraz w sposób niezamierzony przez wielu późniejszych użytkowników cybernetycznej metafory. Cybernetyka badała sterowanie poprzez sprzężenie zwrotne: system obserwuje rezultat, porównuje go z celem i koryguje działanie. Cyfrowe platformy tworzą środowisko, w którym podobna logika zostaje zastosowana nie tylko do maszyn, lecz do interakcji maszyna-człowiek.

Treść jest wyświetlana, reakcja mierzona, model aktualizowany, a kolejna treść selekcjonowana. Użytkownik nie jest biernym elementem tej pętli – sam ją zasila i modyfikuje, mając teoretyczną

możliwość wyjścia z niej. System również nie jest jednolitym podmiotem o spójnej intencji; jego działanie wynika z nakładających się warstw modeli, projektów, celów biznesowych i decyzji organizacyjnych. Strukturalne podobieństwo do sprzężenia zwrotnego pozostaje jednak kluczowe: zachowanie człowieka staje się jednocześnie rezultatem procesu i nową informacją wejściową. To prawdopodobnie najuczciwsza naukowo definicja agency hackingu. Nie chodzi o przejęcie woli, lecz o zbudowanie adaptacyjnego środowiska, które uczy się na podstawie reakcji człowieka, tak jak człowiek uczy się na podstawie reakcji środowiska. Dwie pętle uczenia zaczynają uczyć się siebie nawzajem.

Reranking jako dowód, że architektura ma znaczenie. Badania pokazują, że technologia rekomendacyjna może być projektowana inaczej. Eksperyment opublikowany w „Science" w 2025 roku badał możliwość ograniczania ekspozycji na treści wyrażające partyjną wrogość poprzez zmianę rankingu feedu. Autorzy wykazali, że modyfikacja kryteriów rekomendacji może redukować kontakt z partyjną animozją bez konieczności rezygnacji z całej infrastruktury rekomendacyjnej. Odkrycie to ma ogromną wagę filozoficzną, gdyż przenosi debatę z poziomu demonologii na poziom projektowania instytucjonalnego. Problemem nie jest „algorytm" jako metafizyczne zło, lecz funkcje celu, ograniczenia, kryteria rankingu, metryki sukcesu i mechanizmy odpowiedzialności. Technologia może wzmacniać jedne właściwości środowiska i osłabiać inne, dlatego pytanie polityczne powinno brzmieć nie: „czy używać algorytmów?", lecz: „jakie dobra system ma optymalizować i jakie koszty wolno mu ignorować?".

W języku Arystotelesa powraca telos. Żaden system rekomendacyjny nie jest filozoficznie neutralny, ponieważ musi w jakiś sposób rozstrzygnąć, co oznacza „lepsza" kolejność. Największą różnicę może zatem tworzyć nie moc obliczeniowa modelu, lecz odpowiedź na fundamentalne pytanie: ku jakiemu dobru kierujemy jego sprawność?

Ochrona autonomii przed asymetrią wpływu i brakiem tarcia

Neurorights: kiedy prywatność dociera za czaszkę. Challenger prowadzi swój argument dalej, ku neurorights. Ochrona praw kognitywnych jawi się tu jako odpowiedź na ryzyko technicznego naruszania autonomii, mentalnej prywatności i ciągłości podmiotowości. W 2025 roku problem ten zyskał istotny wymiar instytucjonalny: UNESCO przyjęło Recommendation on the Ethics of Neurotechnology – pierwszy globalny instrument normatywny w tej dziedzinie, podkreślający znaczenie mentalnej prywatności, autonomii, godności oraz ochrony szczególnie wrażliwych danych neuronalnych. Neurotechnologia tworzy bowiem jakościowo nową kategorię problemu.

Tradycyjna prywatność chroniła dom, korespondencję czy tajemnicę komunikowania się; rozwój metod odczytywania i modulowania aktywności układu nerwowego rodzi pytanie, czy ochronie powinna podlegać również wewnętrzna przestrzeń procesów mentalnych, zanim zostaną one wyrażone w zachowaniu lub słowie. UNESCO wiąże prywatność mentalną z tożsamością osobową i sprawstwem, a etyka neurotechnologii coraz częściej traktuje autonomię jako zagadnienie wykraczające poza klasyczną zgodę na procedurę medyczną.

Tu filozofia liberalna powraca do swojego najbardziej pierwotnego pytania. Kant wiązał wolność z autonomią rozumnego podmiotu, Mill bronił przestrzeni indywidualnego rozwoju, a współczesne rozumienie prywatności obejmuje sferę decyzji niezależną od arbitralnego wpływu zewnętrznego. Neurotechnologia wymusza jednak przesunięcie granicy: nie wystarczy pytać, czy ktoś może swobodnie wypowiedzieć przekonanie, jeśli technologia zaczyna ingerować w procesy, z których to przekonanie, nastrój albo impuls się wyłaniają. Należy jednak wprowadzić istotną poprawkę wobec argumentacji Challenger: nie każda perswazja jest przemocą. Trzeba tu sprzeciwić się samej autorce – człowiek zawsze funkcjonował w środowisku wpływu. Rodzice kształtują dzieci, nauczyciele uczniów, przyjaciele siebie nawzajem, kaznodzieje wspólnoty, a reklama konsumentów. Demokracja w ogóle nie mogłaby istnieć bez prób przekonywania innych.

Nie możemy zatem zdefiniować autonomii jako stanu całkowitego braku oddziaływań zewnętrznych. Problem leży gdzie indziej: w asymetrii wiedzy, nieprzejrzystości mechanizmów i zdolności do adaptacyjnego profilowania wpływu. Rozmówca próbujący mnie przekonać dysponuje zazwyczaj ograniczoną wiedzą o mojej historii reakcji. System oparty na danych behawioralnych może natomiast operować znacznie większą liczbą obserwacji, testować niezliczone warianty bodźców i optymalizować kolejność komunikatów, nie ujawniając użytkownikowi kryteriów tej optymalizacji.

Wolność nie wymaga świata wolnego od perswazji. Wymaga jednak warunków, w których podmiot zachowuje zdolność do rozpoznania, refleksji, sprzeciwu i zmiany środowiska, zamiast stawać się wyłącznie obiektem adaptacyjnego eksperymentu. Podmiotowość potrzebuje tarcia. Humanistyka może tu zaproponować pojęcie, którego Challenger nie formułuje wprost: tarcie poznawcze. Świat całkowicie dopasowany do chwilowych preferencji człowieka nie musi być światem najlepiej służącym jego autonomii. Uczenie wymaga spotkania z czymś nieoczekiwanym, demokracja – z argumentem przeciwnika, nauka – z falsyfikującym wynikiem, etyka – z roszczeniem drugiego człowieka, a dojrzewanie – z frustracją, której nie można natychmiast usunąć.

System rekomendacyjny zaprojektowany wyłącznie wokół przewidywalnej satysfakcji może usuwać część tego tarcia, choć żadna platforma nie robi tego całkowicie. Eksperymenty pokazują, że zwiększenie ekspozycji na odmienne treści nie musi automatycznie zmniejszać polaryzacji; samo

„pokazanie drugiej strony” nie zastępuje instytucji rozmowy i wspólnych norm. Hannah Arendt zauważyła, że wolność wymaga pluralności – świata zamieszkanego przez ludzi, którzy nigdy nie będą tacy jak my. Challenger mogłaby dopowiedzieć językiem biologii, że zdrowe psyche nie jest systemem, któremu świat zawsze dostarcza to, czego właśnie zapragnął, lecz organizmem zdolnym utrzymywać siebie pośród różnic i nieprzewidywalności.

"Common sense" jako mechanizm odporności Wspólny Zmysł otrzymuje zatem znaczenie społeczne. Nie jest wyłącznie integrowaniem sygnałów z ciała, lecz zdolnością wspólnoty do utrzymywania rzeczywistości, która nie rozpada się na prywatne strumienie relewancji. Nauka sama w sobie jest przykładem takiej instytucjonalnej technologii wspólnego zmysłu: wymaga metod jawnych, reprodukowalności i procedur korygowania błędów. Można powiedzieć, że metoda naukowa jest kulturową formą allostazy epistemicznej. Nie chroni społeczeństwa przed błędem poprzez zamrożenie jednej prawdy, lecz dzięki mechanizmom korekty pozwala zmieniać przekonania tak, aby utrzymać kontakt z rzeczywistością. Falsyfikacja, recenzja i publiczna krytyka są dla zbiorowej wiedzy tym, czym sprzężenie zwrotne dla organizmu: sposobami wykrywania rozbieżności między modelem a światem. Crazescape jest przeciwieństwem takiej kultury nie dlatego, że zawiera emocje czy konflikt – nauka również je zna – lecz wtedy, gdy środowisko komunikacyjne nagradza transmisję niezależnie od korekty, a szybkość rozchodzenia się komunikatu przewyższa zdolność wspólnoty do jego weryfikacji.

Ochrona psychosfery jako warunek zachowania ludzkiej autonomii

Psychosfera staje się problemem konstytucyjnym. Na tym etapie analizy widać już wyraźnie, dlaczego Challenger przechodzi od biologii do praw kognitywnych. Jeśli sprawstwo jest fundamentalną cechą żywego organizmu, a ludzka wolność stanowi rozwiniętą zdolność organizowania własnego działania poprzez refleksję, pamięć, przewidywanie i relacje społeczne, to infrastruktura ingerująca w warunki tej refleksji przestaje być jedynie narzędziem konsumpcji. Staje się ona częścią środowiska, w którym konstytuuje się obywatel. W klasycznej teorii konstytucyjnej państwo miało ograniczać własną władzę wobec jednostki; w epoce cyfrowej pojawia się jednak pytanie, czy ochrona autonomii nie wymaga również regulowania prywatnych infrastruktur, które organizują ogromną część przestrzeni informacji, uwagi i komunikacji. Nie jest to wezwanie do cenzury czy państwowego sterowania treścią. Wręcz przeciwnie – zdrowy rozsądek polityki liberalnej nakazuje wystrzegać się zarówno koncentracji władzy publicznej, jak i prywatnej.

Istotą problemu jest ustalenie warunków przejrzystości, odpowiedzialności, możliwości wyboru oraz ochrony mentalnej autonomii.

Challenger otwiera tutaj drogę do pojęcia psychosfery w jej najdojrzalszej postaci. Psychosfera nie będzie jedynie poetyckim określeniem planety pełnej czujących istot, lecz przestrzenią, w której ucieleśnione podmioty wytwarzają własne znaczenia, a jednocześnie obszarem, który człowiek zaczął technicznie przeprojektowywać z niespotykaną dotąd precyzją. Autor ujmuje jej ochronę jako obronę biologicznej i kognitywnej wolności przed przekształceniem żywych sprawców w przewidywalne elementy systemu. W tym punkcie biologiczne pytanie o psyche styka się z najstarszym dylematem politycznym: kto ma władzę nad warunkami, w których człowiek staje się sobą? Foucault wskazałby tu na konieczność analizy technik kierowania zachowaniem; Arendt przypomniałaby o wspólnym świecie i pluralności; nauki kognitywne obnażą ograniczenia uwagi i podatność na uczenie nagrodowe, zaś neuroetyka zaczyna budować język ochrony sfery mentalnej.

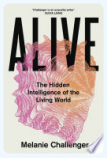
Tym samym agency hacking nie powinien być rozumiany jako fantastyczna opowieść o algorytmie przejmującym wolną wolę. Jest zjawiskiem znacznie bardziej realistycznym i przez to poważniejszym: możliwością przemysłowego projektowania środowiska bodźców w taki sposób, aby coraz skuteczniej przewidywać i zmieniać prawdopodobieństwo ludzkiego działania. Człowiek pozostaje podmiotem, lecz jego podmiotowość nie rozgrywa się już wyłącznie między nim a światem natury; coraz częściej pojawia się trzeci uczestnik – techniczna infrastruktura ucząca się z reakcji żywego organizmu.

Prowadzi to do ostatniego, wielkiego rozszerzenia perspektywy Challengera. Skoro każda psyche tworzy świat znaczeń, organizmy zmieniają fizyczną Ziemię, ludzie wytwarzają technosferę, a ta zwrotnie przekształca warunki ludzkiej i pozaludzkiej sprawczości, należy zapytać, czym stała się planeta jako całość. Nie Gają w sensie jednego superorganizmu, nie wyłącznie Wiernadskiańską biosferą czy Teilhardowską noosferą rozumu, lecz czymś, co Challenger nazywa sferą niezliczonych ucieleśnionych punktów widzenia. Następnym krokiem będzie zatem Psychosfera: Wiernadsky, Teilhard de Chardin, technosfera Petera Haffa, mineralogiczna historia Roberta Hazena, krytyka Gai i planetarna moc metabolizmu. Pojawi się także najtrudniejsza etycznie część „Alive” – pytanie, czy bioinżynieria i CRISPR są jedynie kolejnymi narzędziami człowieka, czy mogą stać się ingerencją w samą autonomię żywych form.

> W świecie, w którym maszyny perfekcyjnie imitują ludzkie cele, największym wyzwaniem przestaje być pytanie o to, czy AI może stać się świadoma. Prawdziwym dramatem jest ryzyko, że my – w pogoni za cyfrową optymalizacją i wygodą – zaczniemy przypominać maszyny: sprawne, przewidywalne, lecz pozbawione własnego „dlaczego”.

> Czy w rzeczywistości zaprojektowanej tak, by usuwać każde poznawcze tarcie, pozostanie jeszcze miejsce na wolność, która z definicji wymaga oporu i nieprzewidywalności?

Inspiracje

[1] 

"Alive" Melanie Challenger

2026 | Canongate Books | ISBN: 9781805300052

Melanie Challenger jest brytyjską pisarką, badaczką i prezenterką radiową, specjalizującą się w historii środowiska, filozofii nauki oraz relacjach między ludźmi a światem ożywionym. Absolwentka Exeter College w Oksfordzie, rozpoczęła karierę w sztukach kreatywnych jako poetka i librecistka, zanim przeszła do literatury faktu. W swojej twórczości zgłębia tematykę natury ludzkiej, świadomości zwierząt i etyki ekologicznej. Challenger pełni ważne funkcje kierownicze, będąc wiceprzewodniczącą Rady ds. Bioetyki Nuffield i wiceprezeską RSPCA. Jest również członkinią-założycielką projektu Animals in the Room, którego celem jest włączenie interesów zwierząt innych niż człowiek do kluczowych procesów decyzyjnych. Jej twórczość, w tym dzieła takie jak „Jak być zwierzęciem" i „Alive", często opowiada się za bardziej zniuansowanym, ucieleśnionym rozumieniem inteligencji i naszego miejsca w świecie przyrody.

Melanie Challenger *is a British writer, researcher, and broadcaster specializing in environmental history, the philosophy of science, and the relationship between humans and the living world. An alumna of Exeter College, Oxford, she began her career in the creative arts as a poet and librettist before transitioning to non-fiction. Her work explores themes of human nature, animal sentience, and ecological ethics. Challenger holds significant leadership roles, serving as the Deputy Co-Chair of the Nuffield Council on Bioethics and a Vice President of the RSPCA. She is also a founding member of the Animals in the Room project, which seeks to integrate the interests of non-human animals into critical decision-making processes. Her writing, including works like 'How to Be Animal' and 'Alive', often advocates for a more nuanced, embodied understanding of intelligence and our place within the natural world.*

Nuffield Council on Bioethics; RSPCA

Literatura uzupełniająca

Książki

Książki autora i dzieła przywołane w tekście:



[1] "How to Be Animal"

Melanie Challenger

2021 | Canongate Books | ISBN: 9781786895745



[2] "On Extinction"

Melanie Challenger

2021 | Catapult | ISBN: 9781640094635



[3] "Galatea"

Melanie Challenger

2006

[4] "The Tender Map"

Melanie Challenger

2016



[5] "Noi, animali"

Melanie Challenger

2023 | Edizioni Sonda | ISBN: 9788872249857

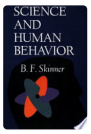


[6] "Wir Tiere"

Melanie Challenger

2021 | btb Verlag | ISBN: 9783641248970

Książki autorów przywoływanych i literatura pokrewna:



[7] "Science and Human Behavior"

B.F. Skinner

Simon and Schuster | ISBN: 9780029290408

[8] "Science and Human Behavior (ʿĀlam al-maʿrifah, 1980)"

B.F. Skinner

[9] "The human phenomenon"

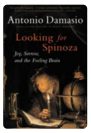
Pierre Teilhard De Chardin



[10] "Being You"

Anil Seth

Faber & Faber | ISBN: 9780571337736



[11] "Looking for Spinoza: Joy, Sorrow, and the Feeling Brain"

Antonio Damasio

Houghton Mifflin Harcourt | ISBN: 9780156028714



[12] "Descartes' error : emotion, reason, and the human brain"

Antonio Damasio

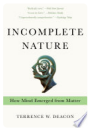
Penguin | ISBN: 9781101640135



[13] "The Music of Life: Biology Beyond the Genes"

Denis Noble

OUP Oxford | ISBN: 9780199295739



[14] "Incomplete Nature"

Terrence Deacon

2011 | W. W. Norton & Company | ISBN: 9780393080834



[15] "The Symbolic Species"

Terrence Deacon

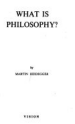
Springer Science & Business Media | ISBN: 9789400723351



[16] "The World of Perception"

Maurice Merleau-Ponty

2020 | Routledge | ISBN: 9781000154900



[17] "What is Philosophy?"

Martin Heidegger



[18] "The question concerning technology and other essays"

Martin Heidegger

Facsimiles-Garl

[19] "Cours sur la Perception (1964-1965)"

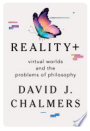
Gilbert Simondon

Editions de la Transparence | ISBN: 9782350510125

[20] **"Deux leçons sur l'animal et l'homme"**

Gilbert Simondon

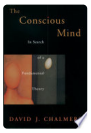
ISBN: 9782340037793



[21] **"Reality+"**

David Chalmers

W. W. Norton & Company | ISBN: 9780393635812



[22] **"The Conscious Mind"**

David Chalmers

Oxford Paperbacks | ISBN: 9780195117899

[23] **"The Portable Hannah Arendt"**

Hannah Arendt



[24] **"The Jewish Writings"**

Hannah Arendt

Schocken | ISBN: 9780307496287



[25] **"The Promise of Politics"**

Hannah Arendt

Schocken | ISBN: 9780307542878

[26] **"The technological Bluff"**

Jacques Ellul

ISBN: 9780802836786

[27] **"The Technological Society"**

Jacques Ellul

1967



[28] **"The Technological System"**

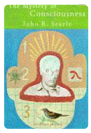
Jacques Ellul

2018 | Wipf and Stock Publishers | ISBN: 9781532615252

[29] **"The rational and social Foundations of music"**

Max Weber

ISBN: 9780758136602



[30] "The Mystery of Consciousness"

John Searle

ISBN: 9781862070745



[31] "The Construction of Social Reality"

John Searle

Simon and Schuster | ISBN: 9781439108369



[32] "John Searle: Our shared condition -- consciousness"

John Searle

Cambridge University Press | ISBN: 9780521797047



[33] "Alterity and Transcendence"

Emmanuel Levinas

Columbia University Press | ISBN: 9780231116510



[34] "Otherwise than Being"

Emmanuel Levinas

Springer Science & Business Media | ISBN: 9789024723744



[35] "Totality and Infinity"

Emmanuel Levinas

ISBN: 9780820702452

[36] "Escape from Freedom"

Erich Fromm

1993 | Harcourt Brace College Publishers | ISBN: 9780030084553



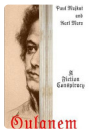
[37] "The Anatomy of Human Destructiveness"

Erich Fromm

New York : Holt, Rinehart and Winston

[38] "A Protest against Law-Taxes"

Jeremy Bentham



[39] "Oulanem"

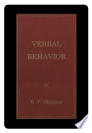
Karol Marks

ISBN: 9780615959931

[40] "The Holy Family"

Karol Marks

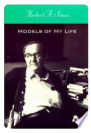
ISBN: 9781983052088



[41] "Verbal Behavior"

B.F. Skinner

B. F. Skinner Foundation | ISBN: 9780989983907



[42] "Models of my life"

Herbert Simon

MIT Press | ISBN: 9780262326582

[43] "Human Problem Solving"

Herbert Simon

[44] "Life of Jesus"

Georg Wilhelm Friedrich Hegel



[45] "Hegel and the Human Spirit"

Georg Wilhelm Friedrich Hegel, Leo Rauch

1983

[46] "Mind, Self and Society"

George Herbert Mead

[47] "The Individual and the Social Self"

George Herbert Mead

1982 | ISBN: 9780226516738



[48] "George Herbert Mead on Social Psychology"

George Herbert Mead

1964

[49] "The Discourses"

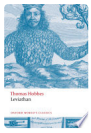
Niccolò Machiavelli

[50] "The Portable Machiavelli"

Niccolò Machiavelli

[51] "Fyrsten"

Niccolò Machiavelli



[52] "Leviathan"

Thomas Hobbes

2024 | Oxford University Press | ISBN: 9780198872474

[53] "Behemoth"

Thomas Hobbes

[54] "De Cive"

Thomas Hobbes

[55] "Invention: The Care and Feeding of Ideas"

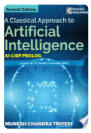
Norbert Wiener

[56] "In memoriam N. M. Sibirtsev"

Vladimir Vernadsky

[57] "La biosphère / Wladimir Vernadsky. - 1929"

Vladimir Vernadsky



[58] "A Classical Approach to Artificial Intelligence"

Munesh Chandra Trivedi

KHANNA PUBLISHING HOUSE | ISBN: 9788190698894

Artykuły naukowe

Artykuły naukowe autora:

[1] "Mental slander"

Melanie Challenger

2021 | New Scientist | DOI: 10.1016/s0262-4079(21)00324-9

Źródło

[2] "The 'Stolen Voices' project for the United Nations International Day of Peace, Imperial War Museum, London"

Melanie Challenger

2007 | Intercultural Education | DOI: 10.1080/14675980701768226

Źródło

[3] "Reassessing the Three Rs?"

Melanie Challenger

2020 | Hastings Center Report | DOI: 10.1002/hast.1143

Źródło

[4] "10 Careful Dignity: A Promising Foundation for Our Relations with Nonhuman Animals"

Melanie Challenger

2026 | Feminist Dream Labs | DOI: 10.59962/9780774872737-012

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: e85f2e8b-4f8b-4ae2-aefd-01afe2e6f5c1