

# Autonomia i tożsamość w dobie neurotechnologii oraz AI: perspektywa Masuda Husaina i Our Brains Our Selves

---



**AUTOR**

Fundacja Dobre Państwo

## STRESZCZENIE / ABSTRACT

Artykuł analizuje ryzyka związane z delegowaniem funkcji poznawczych do systemów AI i neurotechnologii. Autor zwraca uwagę na zjawisko automation bias oraz niebezpieczeństwa związane z offloadingiem, które mogą prowadzić do utraty rezyliencji poznawczej i degradacji biologicznych kompetencji człowieka. Szczególną uwagę poświęcono protezom pamięci, wskazując, że ingerencja w procesy zapamiętywania wpływa bezpośrednio na tożsamość narracyjną i autobiografię jednostki. Tekst podkreśla, że technologia nie jest neutralnym narzędziem, lecz mediatorem współkształtującym podmiotowość. Wnioskiem jest konieczność zachowania metapoznawczej zdolności oceny wiarygodności systemów oraz dbałość o suwerenność funkcjonalną w relacji człowiek-system, aby optymalizacja techniczna nie odbywała się kosztem autentycznej autonomii i roztropności praktycznej.

The article analyzes the risks associated with delegating cognitive functions to AI systems and neurotechnologies. The author draws attention to the phenomenon of automation bias and the dangers related to offloading, which may lead to a loss of cognitive resilience and the degradation of human biological competencies. Particular attention is paid to memory prosthetics, indicating that interference in memorization processes directly affects an individual's narrative identity and autobiography. The text emphasizes that technology is not a neutral tool but a mediator that co-shapes subjectivity. The conclusion is the necessity of maintaining the metacognitive ability to assess system reliability and ensuring functional sovereignty in the human-system relationship, so that technical optimization does not occur at the expense of authentic autonomy and practical wisdom.

Artykuł analizuje krytyczny punkt styku między ludzką podmiotowością a postępującą integracją z neurotechnologiami i sztuczną inteligencją. Autor stawia tezę, że największym zagrożeniem nie jest sama techniczna ingerencja w mózg, lecz ryzyko utraty suwerenności poznawczej – sytuacji, w której zwiększenie funkcjonalnej wydajności człowieka odbywa się kosztem jego zdolności do samodzielnego wyznaczania celów i wartości. Tekst bada zjawiska takie jak 'automation bias', delegowanie funkcji pamięciowych oraz wpływ systemów adaptacyjnych na tożsamość narracyjną. Kluczowym wnioskiem jest postulat przejścia od etyki zakazującej ingerencji ku teorii warunków podmiotowego korzystania z technologii, gdzie nadrzędnym kryterium staje się zachowanie hierarchicznego sprawstwa: automatyzacja może dotyczyć poziomów wykonawczych, ale kontrola nad sensem i kierunkiem życia musi pozostać w gestii człowieka. W obliczu powstawania 'egzosomatycznej warstwy podmiotowości', artykuł wzywa do redefinicji praw człowieka w stronę ochrony integralności mentalnej i prawa do otwartej przyszłości.

## SŁOWA KLUCZOWE

neurotechnologia

sztuczna inteligencja

autonomia relacyjna

suwerenność funkcjonalna

neurorights

automation bias

offloading poznawczy

tożsamość narracyjna

Value Sensitive Design

interfejs mózg-komputer BCI

prywatność mentalna

integralność psychiczna

mediacja technologiczna

neurosecurity

dynamic consent

## Ryzyko utraty czujności i tożsamości w procesie delegowania funkcji poznawczych

W teorii epistemicznej pojawia się problem automation bias. Jeśli system techniczny zazwyczaj okazuje się skuteczny, użytkownik może nadmiernie obniżyć czujność wobec jego błędów. Zjawisko to nie jest nową cechą AI; badano je od dawna w systemach wspomagających lotnictwo, medycynę czy kontrolę procesów. Generatywna AI utrudnia jednak sprawę, gdyż jej odpowiedzi mogą być językowo przekonujące niezależnie od ich poprawności epistemicznej. Z punktu widzenia podmiotowości kluczowy staje się zatem nie tylko dostęp do zewnętrznej inteligencji, ale zachowanie metapoznawczej zdolności oceniania jej wiarygodności. Można to opisać za pomocą

teorii epistemic vigilance Dana Sperbera i jego współpracowników. Ludzie są gatunkiem silnie zależnym od informacji przekazywanych społecznie, dlatego muszą dysponować mechanizmami oceny źródła i treści. Zaufanie nie jest racjonalne wtedy, gdy jest maksymalne, lecz gdy pozostaje odpowiednio skalibrowane.

W środowisku AI funkcja ta staje się trudniejsza, ponieważ jedno narzędzie może agregować setki anonimowych źródeł i przedstawiać syntezę językiem pozbawionym łatwo dostrzegalnych znaczników epistemicznego pochodzenia. Przyszłe rozszerzenie poznawcze może zatem zwiększać kompetencję systemu człowiek-AI, a równocześnie ograniczać określone kompetencje człowieka pozostającego bez tego wsparcia. Nie jest to logicznie sprzeczne. Pilot korzystający z automatyzacji może bezpieczniej prowadzić statek powietrzny, a zarazem rzadziej ćwiczyć manualne umiejętności. Podobnie użytkownik zewnętrznej pamięci może zapamiętywać więcej w sensie funkcjonalnym, lecz mniej w sensie biologicznego przechowywania danych.

Badania nad offloadingiem wskazują właśnie na ten dualizm: korzyści oraz koszt zależności. Pojawia się zatem kwestia rezyliencji poznawczej. Jeśli narzędzie staje się częścią rutynowego procesu działania, awaria systemu może być równoznaczna z nagłą utratą funkcji. Im większy zakres kompetencji delegujemy na zewnątrz, tym bardziej autonomia wymaga nie tylko posiadania narzędzia, lecz zapewnienia trwałości dostępu, możliwości migracji danych oraz alternatywnych procedur w sytuacji odłączenia. Cyberbezpieczeństwo staje się w tym momencie częścią neuroetyki, ponieważ bezpieczeństwo zewnętrznego komponentu może warunkować sprawność całego systemu poznawczego. Jeszcze bardziej radykalny wariant pojawia się przy protezie pamięci hipokampalnej. Badania eksperymentalne próbują modelować transformacje neuronalne zachodzące pomiędzy polami hipokampa i wykorzystywać stymulację do wspierania kodowania informacji. Praca z 2024 roku dotycząca rozwoju hipokampalnej protezy neuronalnej opisywała zastosowania modeli stymulacji mających ułatwiać kodowanie i przywoływanie określonych cech oraz kategorii informacji. Dziedzina ta pozostaje daleka od „wgrywania wspomnień” w popularnym sensie, ale sam kierunek badań odsłania istotny problem: jeżeli technologia może wpływać na to, co zostanie zapamiętane, ingeruje w jeden z mechanizmów tworzących przyszłą autobiografię. Pamięć nie jest bowiem jedynie usługą poznawczą – jest mechanizmem tożsamościowym.

Zwiększenie trafności zapamiętywania może sprawiać wrażenie czystego ulepszenia, jednak selekcja pamięci zawsze niesie konsekwencje dla obrazu przeszłości, który zostaje utrwalony. Zdrowy system nie pamięta wszystkiego. Zapominanie umożliwia generalizację, aktualizowanie przekonań, osłabienie pewnych śladów emocjonalnych i redukcję interferencji. Technologia pamięci musi więc rozwiązać paradoks: optymalizacja retencji nie jest automatycznie optymalizacją podmiotowości. W tym punkcie neurotechnologia spotyka się z filozofią pamięci Paula Ricoeura.

Pamięć jest dla tożsamości ważna nie dlatego, że dostarcza kompletnego archiwum, lecz ponieważ umożliwia organizację przeszłości w znaczącą historię. Absolutny zapis życia mógłby paradoksalnie utrudniać narracyjne selekcjonowanie tego, co istotne. Technologiczne zwiększenie pamięci może zatem poprawić funkcję archiwalną, niekoniecznie jednak usprawniając zdolność nadawania doświadczeniu sensu.

## Autonomia jako właściwość relacji człowiek-system

Podobnego rozróżnienia należy dokonać w przypadku AI. Dostęp do większej liczby informacji nie jest tożsamy z wiedzą, a zdolność wygenerowania wielu argumentów nie oznacza posiadania osądu. Arystotelesowska **phronesis**, czyli roztropność praktyczna, odnosiła się do umiejętności właściwego działania w konkretnej sytuacji, w której ogólna reguła okazuje się niewystarczająca. System sztucznej inteligencji może rozszerzyć informacyjny horyzont decyzji, jednak samo zwiększenie liczby dostępnych opcji nie rozstrzyga, które z nich są zgodne z wartościami danej osoby.

W filozofii techniki problem ten można opisać za pomocą pojęcia **mediacji technologicznej** Petera-Paula Verbeeka. Technologie nie są neutralnymi narzędziami umieszczonymi między podmiotem a światem; one współkształtują sposób, w jaki postrzegamy rzeczywistość i jakie działania uznajemy za możliwe. Termometr przekształca subiektywne odczucie ciepła w liczbę, GPS reprezentuje miasto jako sieć zoptymalizowanych tras, a algorytm rekomendacyjny przedstawia kulturę jako uporządkowany zbiór przewidywanych preferencji. AI może analogicznie zmieniać epistemiczną strukturę rzeczywistości, ponieważ określa, co pojawia się przed człowiekiem jako pierwsze, typowe, prawdopodobne lub godne uwagi. Nie prowadzi to jednak do technologicznego determinizmu.

Użytkownicy reinterpretują narzędzia, instytucje wprowadzają regulacje, a technologie są projektowane w oparciu o różne wartości. Właśnie dlatego **Value Sensitive Design** proponuje uwzględnianie takich wartości jak autonomia, prywatność, sprawiedliwość czy kontrola już na etapie projektowania systemu, a nie dopiero po jego wdrożeniu. Neurotechnologia szczególnie wymaga takiego podejścia, gdyż późniejsza „korekta etyczna” może okazać się trudna, gdy konkretna architektura zostanie głęboko zintegrowana biologicznie i społecznie. W systemach BCI oznacza to na przykład zaprojektowanie wyraźnego kanału zatwierdzania komunikatu, możliwości przerwania dekodowania, informacji o poziomie niepewności algorytmu oraz bezpiecznego trybu awaryjnego. W adaptacyjnej DBS będzie to możliwość kontroli parametrów, śledzenia zmian w działaniu urządzenia oraz odróżnienia decyzji klinicznych od automatycznych adaptacji.

W przypadku AI wspomagającej decyzje kluczowe jest ujawnianie pochodzenia informacji, umożliwienie porównania alternatyw i zachowanie przestrzeni na odmowę. Autonomia nie jest bowiem wyłącznie wewnętrzną cechą użytkownika; stanowi ona również **właściwość relacji człowiek-system**, którą można technicznie wzmacniać lub ograniczać. Tak rozumiana autonomia zbliża się do koncepcji **autonomii relacyjnej** rozwijanej w bioetyce i filozofii feministycznej. Tradycyjny model liberalny przedstawiał autonomiczną osobę jako niezależnego decydenta, chronionego przede wszystkim przed ingerencją z zewnątrz. Ujęcia relacyjne wskazują natomiast, że zdolność do autonomii rozwija się dzięki językowi, edukacji, opiece, relacjom zaufania i instytucjom. Człowiek nie przestaje być autonomiczny dlatego, że korzysta z pomocy. Pytanie brzmi, czy pomoc zwiększa jego zdolność do kierowania własnym życiem, czy przejmuje za niego definiowanie celów.

Ten punkt jest zasadniczy w kontekście osób z niepełnosprawnością. Neuroproteza mowy może sprawić, że człowiek wcześniej pozbawiony skutecznego kanału komunikacji ponownie będzie zdolny do wyrażania preferencji, uczestniczenia w relacjach i odrzucania propozycji. Badania nad neuroprotezami mowy należy zatem rozumieć nie tylko jako rozwój inżynierii sygnałowej, ale jako technologię **przywracania społecznego sprawstwa**. Jest to istotna korekta wobec lęku, że każda integracja człowieka z maszyną musi odbierać autonomię. Czasem zachodzi proces odwrotny: bez urządzenia człowiek posiada intencje, lecz nie może ich skutecznie wyrazić. Neuroetyka powinna zatem rozróżniać **autonomię przez niezależność** od **autonomii przez skuteczne wsparcie**. Ian Waterman odzyskał część możliwości działania dzięki świadomej kontroli wzrokowej, pacjent z afazją może korzystać z urządzenia komunikacyjnego, a osoba z paralizem z BCI. W każdym z tych przypadków system rozszerza relację sprawczą pomiędzy intencją a światem.

## **Technologiczna      zależność      konstytutywna      jako zagrożenie dla sprawstwa**

Zależność od narzędzia nie musi być przeciwieństwem autonomii – podobnie jak zależność od okularów nie odbiera człowiekowi sprawstwa w tym, co czyta. Niemniej jednak taka relacja tworzy nowy rodzaj podatności na władzę. Jeśli kluczowy komponent rozszerzonej podmiotowości znajduje się pod kontrolą przedsiębiorstwa, szpitala lub platformy, pojawiają się fundamentalne pytania o aktualizację, prawo dostępu, ciągłość serwisu, własność danych czy możliwość przeniesienia systemu do innego dostawcy. O ile w przypadku tradycyjnej protezy materialnej problem ten jest istotny, o tyle przy interfejsie głęboko sprzężonym z mową, pamięcią lub regulacją

emocji nabiera on krytycznego znaczenia. W tym miejscu ekonomia platformowa spotyka się z filozofią osoby, a autor wprowadza pojęcie technologicznej zależności konstytutywnej.

Jeśli urządzenie służy jedynie wygodzie, jego utrata obniża komfort życia. Jeśli natomiast staje się warunkiem mowy, ruchu lub stabilnej kontroli objawów, decyzje dostawcy mogą bezpośrednio wpływać na zakres podstawowego sprawstwa użytkownika. Z tej perspektywy prawo do serwisu, interoperacyjności i bezpieczeństwa przestaje być jedynie konsumenckim problemem technicznym; staje się elementem ochrony integralności funkcjonalnej osoby. Szczególnie wyraźnie widać to w kontekście cyberbezpieczeństwa. Implantowany interfejs nie jest wyłącznie urządzeniem medycznym, jeśli jego oprogramowanie wymaga aktualizacji i komunikuje się z zewnętrznymi systemami – staje się on częścią cyberfizycznej infrastruktury człowieka. Wówczas koncepcje cyberbezpieczeństwa (poufność, integralność i dostępność danych) zyskują wymiar antropologiczny. Naruszenie poufności może odsłonić neurodane, naruszenie integralności może zmienić działanie systemu, a utrata dostępności może pozbawić użytkownika funkcji, na której opiera on swoją codzienną autonomię.

W tym punkcie pojęcie „neurosecurity” przestaje być futurystycznym ozdobnikiem. Jeżeli neurotechnologia staje się elementem regulacji biologicznej i sprawstwa, bezpieczeństwo jej infrastruktury staje się jednym z warunków bezpieczeństwa samej osoby. UNESCO podkreśla właśnie ryzyka związane z możliwością dostępu do aktywności mózgu i jej modyfikowania, zwłaszcza w połączeniu z AI.

W międzynarodowej debacie pojawiają się zatem postulaty szczególnej ochrony prywatności mentalnej, integralności psychicznej, wolności myśli oraz ciągłości tożsamości. Pojęcie neurorights pozostaje jednak przedmiotem dyskusji. Jedno stanowisko zakłada, że neurotechnologie generują tak nowe rodzaje zagrożeń, iż konieczne jest ustanowienie nowych praw podstawowych. Inne sugeruje, że istniejące prawa do prywatności, integralności osobistej, wolności myśli i ochrony danych mogą zostać odpowiednio rozwinięte bez mnożenia kategorii prawnych. UNESCO przyjęło kierunek traktujący mental integrity, human dignity, autonomy oraz mental privacy jako centralne wartości wymagające szczególnej ochrony w obliczu rozwoju neurotechnologii.

## **Rozproszenie odpowiedzialności w systemach adaptacyjnych**

Spór ma więc charakter bardziej instytucjonalny niż aksjologiczny: stosunkowo szeroki konsensus dotyczy tego, co wymaga ochrony; znacznie mniejszy zaś tego, jakiej dokładnie formy prawnej ta

ochrona potrzebuje. Jeszcze trudniejsza okazuje się kwestia ciągłości odpowiedzialności w systemach adaptacyjnych. Jeśli urządzenie automatycznie zmienia własne parametry, a zmiana ta wpływa na zachowanie pacjenta, nie wystarczy wskazać jednego podmiotu odpowiedzialnego. Pacjent zgodził się na implantację, klinicysta ustalił parametry, producent zaprojektował algorytm, a system zareagował na dynamiczny stan neuronalny. Prawo zna podobne problemy w przypadku urządzeń automatycznych, jednak w neurotechnologii rezultat może dotyczyć bezpośrednio zachowania i motywacji użytkownika. Rozwiązaniem nie powinno być stwierdzenie, że "algorytm podjął decyzję", gdyż algorytm nie zyskuje statusu moralnego osoby tylko dlatego, że wykonuje selekcję techniczną.

Bardziej użyteczna jest analiza łańcucha kontroli i przewidywalności. Kto określił funkcję działania systemu? Kto mógł przewidzieć ryzyko? Kto miał możliwość interwencji? Czy pacjent znał charakter automatycznej adaptacji? Czy klinicysta mógł wykryć zmianę? Czy producent zapewnił mechanizm bezpiecznej korekty?

Odpowiedzialność w systemie rozproszonym może być wielopodmiotowa bez stawiania się bezpodmiotową. Prowadzi to do rozróżnienia pomiędzy causal responsibility a moral responsibility. Urządzenie może być jednym z czynników przyczynowych działania, ale nie wynika z tego jego odpowiedzialność moralna. Człowiek może stanowić ogniwo w łańcuchu przyczynowym, a równocześnie ponosić ograniczoną odpowiedzialność, jeśli jego zdolność kontroli została technicznie osłabiona. Konstruktor może nie uczestniczyć w konkretnym zdarzeniu fizycznie, a mimo to odpowiadać zawodowo za przewidywalny błąd projektu. Neurotechnologia wyostreza więc problem znany już z teorii organizacji: przyczynowość i odpowiedzialność nie mają tej samej mapy. Szczególne znaczenie uzyskuje także świadoma zgoda. W klasycznym modelu zgoda poprzedza zabieg, lecz w technologii długotrwale adaptacyjnej może to być niewystarczające. Jeśli system z czasem zmienia sposób działania lub jego funkcje są rozszerzane programowo, bardziej adekwatnym rozwiązaniem staje się dynamic consent – zgoda traktowana jako proces, a nie jednorazowy podpis. Pacjent powinien mieć możliwość uzyskiwania informacji o istotnych zmianach oraz ponownej oceny akceptowalności działania systemu. Problem komplikuje się jeszcze bardziej, gdy sama technologia zaczyna wpływać na zdolność udzielania tej zgody.

## Tożsamość narracyjna i prawo do samokształtowania

Terapia neuropsychiatryczna może zmieniać nastrój, impulsywność lub ocenę ryzyka. Pojawia się wówczas pytanie o wymiar czasowy: która wersja preferencji powinna być decydująca – ta sprzed terapii, w trakcie procesu, czy po osiągnięciu stabilnego efektu? Uniwersalnej odpowiedzi nie ma.

Konieczny jest model uwzględniający wcześniejsze wartości, aktualny dobrostan, zdolność rozumienia oraz trwałość nowego stanu. To właśnie tutaj Ricoeurowska tożsamość narracyjna okazuje się bardziej użyteczna niż poszukiwanie statycznego „prawdziwego Ja”. Człowiek może przejść przemianę i nadal pozostawać sobą, o ile potrafi włączyć tę zmianę do własnej historii jako sensowną kontynuację. Interwencja staje się problemem nie dlatego, że powoduje zmianę, ale gdy rozrywa możliwość rozpoznania własnych działań, relacji i celów jako należących do jednej biografii. Autentyczność można zatem rozumieć diachronicznie: nie jako zachowanie każdej wcześniejszej cechy, lecz jako zdolność do refleksyjnego przyswojenia zmiany.

Derek Parfit poszedł jeszcze dalej, kwestionując, czy ścisła tożsamość osobowa jest w ogóle tym, co moralnie najważniejsze. Jego redukcjonistyczne ujęcie akcentowało ciągłość psychologiczną i powiązania między kolejnymi stanami osoby. W kontekście neurotechnologii oznaczałoby to pytanie nie tyle o to, czy po interwencji jesteśmy dokładnie tym samym metafizycznym podmiotem, lecz jak wiele relacji pamięci, intencji, wartości i charakteru łączy stan sprzed i po zabiegu. Perspektywa Parfita jest radykalna, ale pomocna, gdyż przenosi uwagę z nieuchwytnego metafizycznego rdzenia ku mierzalnym i opisywalnym formom ciągłości. Nie rozwiązuje ona jednak problemu normatywnego.

Możliwe jest przecież, że technologia znacząco zmienia człowieka, a on sam ceni ten rezultat. Czy taka zmiana narusza autentyczność? Neuroetyka nie może zakładać, że zachowanie wcześniejszej osobowości zawsze ma pierwszeństwo przed możliwością samotransformacji. Ludzie od zawsze kształtują siebie poprzez edukację, religię, terapię, trening, migrację, relacje czy substancje psychoaktywne. Celowa modyfikacja własnych dyspozycji jest wpisana w historię kultury. Stąd silniejsza teoria autonomii powinna obejmować prawo do samokształtowania, a nie tylko prawo do zachowania status quo. Jednocześnie samokształtowanie wymaga określonych warunków: dostępu do informacji, braku przymusu, możliwości rewizji wyboru oraz ochrony przed trwałą ingerencją dokonaną pod wpływem chwilowego stanu. Neurotechnologia zwiększa moc samomodyfikacji, a wraz z nią ciężar tych wymagań. Granica między terapią a ulepszaniem (enhancement) nie jest zatem najważniejszą linią demarkacyjną w etyce. Ważniejsze stają się pytania o proporcjonalność ryzyka, świadomą zgodę, odwracalność, możliwość korekty, wpływ na osoby trzecie, sprawiedliwy dostęp oraz zdolność użytkownika do zachowania refleksyjnej kontroli. Starsza literatura dotycząca DBS już wskazywała, że standardowe zasady bioetyczne – dobroczynność, nieszkodzenie i autonomia – mogą dostarczać bardziej użytecznej ramy niż abstrakcyjny lęk przed każdą „zmianą osobowości”. Należy do nich dodać czwartą klasyczną zasadę Beauchampa i Childressa: sprawiedliwość.

Jeżeli neurotechnologie będą rozwijać się wyłącznie dla osób z wysokim kapitałem ekonomicznym, ich skutki uderzą w strukturę nierówności. Dostęp do komunikacyjnego BCI u osoby z paraliżem ma charakter medyczny, jednak możliwość wykorzystania przyszłych systemów do zwiększania wydajności zawodowej czy poznawczej może stać się dobrem pozycyjnym. W przypadku dóbr pozycyjnych wartość zależy częściowo od przewagi nad innymi; wtedy technologia nie tylko zwiększa możliwości użytkownika, lecz zmienia względne położenie pozostałych. Socjologia technologii wskazuje ponadto na proces normalizacji. Gdy technologia staje się powszechna, zmienia się standard oczekiwanego działania. Telefon komórkowy był kiedyś udogodnieniem; później stał się praktycznie warunkiem dostępności społecznej i zawodowej.

Analogiczny proces może dotyczyć systemów poznawczego wsparcia. Jeśli AI zacznie powszechnie przyspieszać tempo tworzenia tekstów, analiz i podejmowania decyzji, osoba niekorzystająca z tych narzędzi może zostać oceniona nie według dawnej normy ludzkiej, lecz według nowego standardu człowiek-AI. Michel Foucault nazwałby część tej dynamiki problemem normalizującej władzy: normy nie tylko opisują przeciętność, lecz mogą organizować oczekiwania wobec jednostek. Nie trzeba przyjmować całej foucaultowskiej teorii władzy, aby dostrzec praktyczny problem. Enhancement może przejść od możliwości do oczekiwania, a później od oczekiwania do nieformalnego obowiązku. Gdy organizacja deklaruje, że coś jest „dobrowolne”, ale osoby niekorzystające z narzędzia systematycznie przegrywają w wyścigu o awans, dobrowolność wymaga bardziej rygorystycznej analizy niż tylko braku formalnego nakazu.

Z tego powodu neuroetyka przyszłości będzie musiała współpracować z prawem pracy, ekonomią, teorią organizacji i socjologią, a nie wyłącznie z medycyną. Zdolność do modyfikacji uwagi, pamięci, nastroju i wydajności stanie się problemem instytucjonalnym, jeśli pracodawca uzyska interes ekonomiczny w konkretnym stanie neurokognitywnym pracownika. Granica między terapią a eksploatacją może wówczas zależeć nie od rodzaju urządzenia, lecz od asymetrii władzy i możliwości rzeczywistej odmowy.

## **Ochrona otwartej przyszłości dziecka przed neuroprojektowaniem**

Szczegółnej ochrony wymagają dzieci i młodzież, gdyż ich podmiotowość nie jest jeszcze ustabilizowaną konstrukcją, lecz systemem w procesie rozwoju. UNESCO w debacie o neurotechnologiach podkreśla właśnie kluczowe znaczenie okresu wzrostu mózgu oraz konieczność ochrony wolności i integralności mentalnej młodych ludzi. Problem zgody nabiera tutaj podwójnego wymiaru: ograniczona jest nie tylko obecna zdolność dziecka do zrozumienia sytuacji,

ale sama interwencja może wpłynąć na to, jakim podmiotem stanie się ono w przyszłości. W tym kontekście warto przywołać koncepcję Joela Feinberga – „right to an open future” – czyli prawo dziecka do zachowania szerokiego wachlarza przyszłych możliwości, których nie powinny arbitralnie zamykać decyzje dorosłych.

Neurotechnologia nadaje tej kategorii wyjątkowo konkretny wymiar. O ile interwencja terapeutyczna może tę otwartą przyszłość poszerzać – na przykład przywracając komunikację czy ruch – o tyle enhancement, ukierunkowany na stworzenie określonego profilu poznawczego, budzi pytanie: czy rodzice wspierają zdolności dziecka, czy nadmiernie projektują przyszłą osobę według własnych preferencji? Nie należy jednak zakładać, że tzw. „naturalny rozwój” jest normatywnie neutralny.

Edukacja, wychowanie i trening również głęboko kształtują mózg dziecka. Różnica między szkołą a neurostymulacją nie może zatem sprowadzać się jedynie do tego, czy dana metoda oddziałuje na organ – obie działają na mózg, choć czynią to poprzez zupełnie inne mechanizmy, z różnym stopniem bezpośredniości, ryzyka i możliwości wycofania.

## **Odwracalność techniczna a nieodwracalność biograficzna**

Rzetelna analiza musi uwzględniać stopień inwazyjności, odwracalność, trwałość oraz możliwość świadomego odrzucenia skutków w przyszłości. Pojęcie odwracalności jest jednak bardziej złożone, niż sugeruje język techniczny. Wyłączenie stymulatora nie musi bowiem przywrócić poprzedniego stanu psychologicznego; człowiek w trakcie terapii nabywa nowe doświadczenia, relacje i wspomnienia. Nawet jeśli efekt fizjologiczny jest odwracalny, biografia nie jest. Osoba, która przez pięć lat funkcjonowała dzięki technologii, po jej odłączeniu nie wraca do punktu sprzed implantacji.

Historia sama staje się częścią systemu. Podobnie dzieje się w przypadku AI. Można przestać korzystać z systemu, ale nie da się cofnąć argumentów, decyzji i praktyk poznawczych wypracowanych we współpracy z algorytmem. Technologia pozostawia ślad poprzez proces uczenia. Odwracalność powinna zatem obejmować co najmniej trzy wymiary: fizyczne usunięcie narzędzia, funkcjonalne odzyskanie wcześniejszej zdolności oraz narracyjną możliwość życia bez systemu. Te warstwy mogą się od siebie rozchodzić. W kognitywistyce analogiczny problem opisuje pojęcie scaffolding, czyli rusztowania. Dobre rusztowanie wspiera rozwój kompetencji, a następnie może zostać usunięte. Inne staje się trwałą częścią systemu. Okulary nie mają nas nauczyć widzenia; ich funkcją jest stała kompensacja. Tutor edukacyjny powinien natomiast stopniowo

przekazywać kompetencje uczniowi. Ocena AI wymaga więc pytania, jaki rodzaj rusztowania stworzymy: protezę, partnera, nauczyciela czy substytut kompetencji. Każdy z tych modeli niesie inne konsekwencje dla autonomii. Proteza może stać się permanentnym elementem sprawstwa, nauczyciel powinien zwiększać samodzielność, a partner współdzielić proces decyzyjny. Substytut natomiast może usprawniać wykonanie zadania, jednocześnie osłabiając wewnętrzną zdolność człowieka. Nie można zatem oceniać wszystkich zastosowań AI ogólnikowym hasłem „wspomaga człowieka”.

W tym kontekście autor przywołuje koncepcję Vygotskiego. Jego pojęcie strefy najbliższego rozwoju zakłada, że funkcja najpierw jest wykonywana przy wsparciu bardziej kompetentnego partnera, a później zostaje zinternalizowana. AI mogłoby teoretycznie pełnić podobną funkcję dydaktyczną, gdyby wsparcie było stopniowane i ukierunkowane na rozwój użytkownika. Może jednak działać odwrotnie – wykonywać zadanie tak skutecznie, że proces internalizacji nigdy nie zachodzi. Różnica tkwi nie w samej obecności technologii, lecz w architekturze interakcji i pedagogicznym celu systemu. Prowadzi to do istotnego rozróżnienia między modelami human-in-the-loop, human-on-the-loop oraz praktycznie human-out-of-the-loop. W pierwszym przypadku człowiek aktywnie uczestniczy w każdej krytycznej decyzji. W drugim nadzoruje system, który automatyzuje większość procesu. W trzecim jego udział może stać się ceremonialny. Im bliżej ostatniego wariantu, tym trudniej przypisać człowiekowi pełne autorstwo tylko dlatego, że formalnie uruchomił narzędzie. Jednocześnie nadmierne wymaganie ręcznej kontroli mogłoby zniweczyć korzyści z automatyzacji. Pacjent korzystający z BCI nie powinien sprawdzać każdej milisekundy dekodowania, tak jak zdrowy człowiek nie monitoruje świadomie każdego motoneuronu.

## **Hierarchiczne sprawstwo jako warunek zachowania suwerenności**

Znów powraca kwestia właściwego poziomu kontroli. Autonomia nie wymaga mikrozarządzania każdym etapem procesu; wymaga jednak panowania nad celem, kluczowymi parametrami oraz możliwości przerwania działania. Pozwala to sformułować zasadę hierarchicznego sprawstwa: niższe poziomy kontroli mogą zostać zautomatyzowane, o ile wyższe poziomy celu i normatywnej autoryzacji pozostają po stronie człowieka. Zdrowy układ nerwowy funkcjonuje właśnie w ten sposób – nie kontrolujemy świadomie napięcia każdego mięśnia podczas pisania, lecz świadomie wybieramy to, co chcemy napisać. Technologia staje się problematyczna w momencie, gdy

przesuwa się z poziomu wykonawczego ku poziomowi definiowania celów, omijając kontrolę podmiotu.

Tę samą zasadę można zastosować do oceny AI w procesach intelektualnych. O ile automatyczne formatowanie tekstu niemal nie narusza autorstwa, o tyle zasugerowanie konkretnego argumentu zwiększa udział systemu w procesie epistemicznym. Jeśli system samodzielnie wybiera tezę, argumenty i wnioski, a człowiek jedynie biernie je akceptuje, związek między finalnym tekstem a refleksyjną kontrolą użytkownika ulega znacznemu osłabieniu. Granica ta nie jest ostra, ale można ją analizować poprzez pytanie: który poziom hierarchii decyzyjnej został delegowany? Jest to kluczowe w kontekście odpowiedzialności zawodowej.

Lekarz może wykorzystać AI do wyszukiwania danych, lecz nie może przenieść na system normatywnej odpowiedzialności za decyzję kliniczną tylko dlatego, że rekomendacja została wygenerowana algorytmicznie. Prawnik, naukowiec czy urzędnik również nie stają się mniej odpowiedzialni przez sam fakt, że część procesu poznawczego została wsparta technicznie. Zakres tej odpowiedzialności powinien jednak zależeć od rzeczywistej możliwości kontroli, przewidywalności oraz profesjonalnego obowiązku weryfikacji. W filozofii działania jest to odmiana zasady answerability: sprawca ponosi odpowiedzialność między innymi wtedy, gdy potrafi sensownie odpowiedzieć na pytanie, dlaczego postąpił w dany sposób. Jeśli jedyną odpowiedzią staje się stwierdzenie, że „algorytm tak zasugerował”, odpowiedzialność epistemiczna podmiotu słabnie – nawet jeśli formalna kompetencja do podjęcia decyzji pozostaje przy nim.

Technologia może zatem rozszerzać moc działania szybciej niż zdolność rozumienia podstaw własnych czynów. Problem ten jest istotny dla demokracji wiedzy, nauki i administracji, lecz w wymiarze antropologicznym okazuje się jeszcze bardziej fundamentalny. Być autorem własnego życia nie znaczy generować wszystko samodzielnie; znaczy posiadać wystarczającą zdolność rozpoznania, przyjęcia, odrzucenia i uzasadnienia elementów, które kierują działaniem. Człowiek zawsze korzysta z cudzej wiedzy – autonomia nie wymaga epistemicznej samotności, lecz zachowania możliwości refleksyjnego dystansu wobec źródeł, z których czerpiemy. Właśnie dlatego AI może zarówno wzmacniać, jak i osłabiać autonomię.

## **Rozróżnienie między wydajnością a sprawczością epistemiczną**

Osobie z ograniczeniami językowymi AI może pomóc w komunikowaniu złożonych preferencji, człowiekowi przeciążonemu informacyjnie – uporządkować materiał, a pacjentowi z deficytami

pamięciowymi – przypominać istotne fakty. Ten sam system może jednak zwiększać zależność, maskować brak zrozumienia lub podsuwać odpowiedzi tak płynnie, że użytkownik przestaje dostrzegać różnicę między rozpoznaniem argumentu a jego rzeczywistym pojęciem. Psychologia uczenia opisuje tu zjawisko **illusion of competence**: płynność przetwarzania informacji bywa mylona z opanowaniem wiedzy. Tekst wydaje się zrozumiały, więc czytelnik zakłada, że potrafiłby samodzielnie go odtworzyć lub zastosować; AI generujące klarowne wyjaśnienia może tę pokusę wzmacniać. Jest to problem metapoznawczy, a nie argument przeciwko samemu narzędziu. Rozwiązaniem powinno być projektowanie interakcji wymagającej aktywnego wydobywania, sprawdzania i rekonstrukcji wiedzy – o ile celem jest nauka, a nie tylko jednorazowe wykonanie zadania. Należy zatem odróżnić **epistemic performance** od **epistemic agency**. Pierwsze pojęcie oznacza uzyskanie poprawnej odpowiedzi; drugie zaś zdolność rozumienia, oceny i kontrolowania procesu, który do tej odpowiedzi prowadzi.

Kalkulator radykalnie usprawnia rachunki, ale nie tworzy automatycznie zrozumienia matematyki. Podobnie system AI może przygotować poprawną analizę, podczas gdy użytkownik pozostanie niezdolny do rozpoznania, dlaczego jest ona poprawna. Społeczeństwo wysoko zautomatyzowane może zatem osiągać większą wydajność poznawczą przy nierównomiernym rozkładzie indywidualnej sprawczości epistemicznej. Prowadzi to do teorii **deskilling** – utraty kompetencji wskutek długotrwałej automatyzacji. Socjologia pracy analizowała podobne procesy już od czasów Harry'ego Bravermana: technologia może przenosić wiedzę z pracownika do procedury lub maszyny, co zwiększa wydajność organizacji, a jednocześnie ogranicza indywidualną kontrolę nad procesem pracy. AI przenosi ten problem z obszaru manualnego na czynności językowe i poznawcze. Nie musi on jednak prowadzić do deskillingu w sposób automatyczny. Narzędzie może działać jako **skill amplifier**, jeśli użytkownik zachowuje aktywny udział i wykorzystuje je do eksploracji problemów przekraczających jego wcześniejsze możliwości. Ostateczny skutek zależy więc od instytucji, projektowania interfejsu, edukacji i kultury pracy. To kolejny dowód na to, że konsekwencji technologii nie można wyprowadzać wyłącznie z jej parametrów technicznych.

Ten sam model może być nauczycielem, protezą, współpracownikiem lub substytutem rozumowania – zależnie od sposobu użycia. W perspektywie całej rozprawy prowadzi to do ważnej rewizji kategorii "naturalnego Ja". Człowiek od dawna jest istotą technologiczną: pismo rozszerzyło pamięć, druk skalował wiedzę, zegar zmienił organizację czasu, okulary przebudowały percepcję, a telefon rozszerzył społeczną obecność. Neurotechnologia i AI różnią się jednak głębokością sprzężenia, szybkością adaptacji i zakresem procesów, w które mogą ingerować. Nie tworzą one całkowicie nowego problemu, lecz zaostrzają stary do poziomu, na którym dotychczasowe intuicje przestają wystarczać. Clark określał człowieka jako "natural-born cyborg": organizm niezwykle

zdolny do włączania artefaktów w system poznawczy. Teza ta jest mniej futurystyczna, niż mogłoby się wydawać.

Język sam w sobie jest technologią poznawczą przejętą społecznie; pismo umożliwia operacje, których pamięć biologiczna nie udźwignęłaby na podobną skalę; notacja matematyczna zmienia rodzaj problemów, które człowiek potrafi rozwiązywać. AI stanowi kolejny krok kumulatywnej technologizacji poznania, choć jest to krok jakościowo szczególny, ponieważ zewnętrzne narzędzie samo generuje nowe reprezentacje. Najgłębszym pytaniem nie jest zatem to, czy technologia "wejdzie do Ja" – ona uczestniczy w jego funkcjonowaniu od dawna.

Pytanie brzmi: które poziomy podmiotowości jesteśmy gotowi delegować i pod jakimi warunkami zachowujemy nad nimi normatywne zwierzchnictwo? Możemy powierzyć systemowi zapamiętanie numeru telefonu bez większej utraty autonomii. Delegowanie jednak wyboru przyjaciół, decyzji medycznej, sądu moralnego czy określenia celów życiowych ma inną wagę, gdyż dotyczy coraz wyższych poziomów organizacji osoby. Właśnie tutaj powinna przebiegać jedna z najważniejszych osi przyszłej neuroetyki: nie między tym, co "biologiczne" a „sztuczne”, lecz pomiędzy technologią instrumentalną, która wspiera realizację przyjętych przez człowieka celów, a technologią konstytutywno-normatywną, która zaczyna współkształtować same cele, wartości i kryteria wyboru. Granica ta pozostaje stopniowalna, ale jej znaczenie jest fundamentalne. Nie oznacza to jednak, że drugi typ technologii jest z definicji niedopuszczalny.

## **Autonomia jako proces i model warstwowy**

Psychoterapia, podobnie jak edukacja, literatura czy głęboka relacja, może zmieniać wartości i cele człowieka. Różnica tkwi jednak w transparentności, wzajemności oraz możliwości refleksyjnej odpowiedzi. Terapeuta przedstawia interpretację, z którą pacjent może dyskutować; książka natomiast nie dostosowuje potajemnie każdej tezy do psychologicznego profilu czytelnika.

Algorytm personalizujący oddziaływanie może działać poniżej poziomu świadomej kontroli. Im większa asymetria wiedzy między systemem a użytkownikiem, tym kluczniejsze stają się normy transparentności i zakaz manipulacji. W tym kontekście pojawia się pojęcie autonomii preferencji. Podczas gdy ekonomia tradycyjna często traktowała preferencje jako gotowe dane wejściowe wyboru, psychologia, socjologia i neurobiologia dowodzą, że są one kształtowane przez doświadczenie, kontekst, reklamę, nawyki, stany fizjologiczne oraz uczenie społeczne. Jeśli technologia potrafi przewidywać i modyfikować preferencje, stwierdzenie, że "użytkownik wybrał", przestaje wystarczać. Należy zapytać o proces, w którym powstało to, czego dany człowiek zaczął chcieć.

Nie chodzi tu o powrót do paternalistycznych pytań o to, które pragnienia są "prawdziwe". Trafniejsze jest badanie proceduralne: czy człowiek dysponował informacją i alternatywami, czy miał czas na refleksję i możliwość zmiany decyzji; czy oddziaływanie było transparentne i czy nie wykorzystywało podatności, której osoba nie mogła rozpoznać; czy wreszcie interes systemu był zgodny z interesem użytkownika. Autonomia staje się zatem właściwością procesu powstawania decyzji, a nie tylko faktu naciśnięcia przycisku "zgadzam się".

Z tej perspektywy neural enhancement, generatywna AI i systemy rekomendacyjne należą do jednej rodziny problemów, mimo różnic w biologicznej bezpośredniości. Wszystkie te narzędzia mogą zmieniać przestrzeń dostępnych racji. DBS czyni to poprzez neuromodulację, AI poprzez architekturę informacji, psychofarmakologia poprzez zmianę wartościowania i afektu, a edukacja poprzez modyfikację reprezentacji semantycznych.

Choć nie są one moralnie równoważne, ich porównanie ujawnia, że autonomia nigdy nie była absolutną izolacją od wpływów. Jest ona raczej zdolnością do działania w świecie tych wpływów bez całkowitej utraty refleksyjnej kontroli. Pozwala to sformułować model autonomii warstwowej. Na najniższym poziomie niezbędna jest integralność biologiczna, umożliwiająca świadome doświadczenie. Następnie wymagany jest dostęp do adekwatnej reprezentacji świata, a później zdolność rozumienia znaczenia informacji, zapamiętywania jej i porównywania alternatyw. Kolejną warstwą jest możliwość wartościowania, hamowania i działania, nad którą znajduje się refleksyjna zdolność oceny własnych pragnień i identyfikowania się z nimi. Na najwyższym poziomie leży możliwość osadzenia decyzji w długoterminowej biografii i relacjach społecznych. Technologia może wspierać jedną warstwę, jednocześnie niekorzystnie wpływając na inną, dlatego bilans autonomii nie może być redukowany do pojedynczego wyniku testowego.

BCI może zwiększyć sprawstwo komunikacyjne, nawet jeśli pogłębia zależność technologiczną. DBS może poprawić zdolność działania, niosąc pewne ryzyko zmiany nastroju. AI może rozszerzyć zasób dostępnych argumentów, jednocześnie osłabiając motywację do samodzielnego ich rekonstruowania. Ocena wymaga zatem analizy profilu skutków, a nie prostego werdyktu o "większej lub mniejszej wolności". Podobnie należy myśleć o tożsamości: zmiana charakteru może być pozytywna, negatywna lub normatywnie neutralna. Osoba po skutecznym leczeniu może stwierdzić: "stałam się znowu sobą", podczas gdy inna powie: "jestem kimś nowym i cieszę się z tego".

Badania nad DBS pokazują, że perspektywa samych pacjentów bywa znacznie mniej katastroficzna niż filozoficzne scenariusze o utracie tożsamości. Neuroetyka potrzebuje zatem nie tylko teorii filozoficznych, ale również evidence-based neuroethics – systematycznego włączania doświadczeń rzeczywistych użytkowników technologii. Jest to istotna lekcja metodologiczna: ekspert może uznać

zmianę za naruszenie autentyczności, podczas gdy pacjent odczuwa ją jako odzyskanie życia. I odwrotnie – kliniczna poprawa w skali objawowej może współwystępować z poczuciem obcości wobec siebie. Ani wynik testu, ani introspekcja nie mają monopolu na prawdę. Konieczna jest triangulacja danych: pomiar funkcji, relacja pacjenta, obserwacja bliskich, stabilność efektu w czasie oraz zgodność z wartościami wcześniejszymi i aktualnymi. Tę samą metodę należałoby zastosować do AI – nie wystarczy bowiem mierzyć jedynie wzrostu wydajności.

## **AI jako nowa nisza poznawcza i egzosomatyczna warstwa podmiotowości**

Należy zbadać, czy długotrwałe korzystanie z tych narzędzi zmienia metapoznanie, pamięć, poczucie autorstwa, odporność na utratę narzędzia oraz zdolność do niezależnej oceny informacji. Przegląd badań nad offloadingiem dowodzi już, że wydajność doraźna i trwałość kompetencji to dwie różne zmienne. W przypadku generatywnej AI potrzebujemy analogicznie badań longitudinalnych, a nie wyłącznie porównań jednorazowego wykonania zadania. Z perspektywy teorii ewolucji można wręcz stwierdzić, że człowiek wkroczył w nową fazę konstrukcji niszy poznawczej. Organizmy nie tylko adaptują się do środowiska; niektóre gatunki przekształcają je w taki sposób, że zmienia to presję selekcyjną i warunki dalszego rozwoju.

Człowiek czyni to na wyjątkową skalę za pomocą języka, technologii i instytucji. AI jest narzędziem budowania środowiska, w którym przyszłe pokolenia będą rozwijać uwagę, pamięć i strategie poznawcze w innych warunkach niż ich poprzednicy. Teoria niche construction podważa prostą opozycję biologii i kultury: kultura jest jednym ze środowisk wpływających na rozwój organizmu, a biologicznie wyposażony organizm z kolei tworzy kulturę. Podobna pętla może powstać między ludzkim poznaniem a AI: ludzie tworzą modele w oparciu o wytworzoną przez siebie kulturę, modele te wpływają na kolejne teksty i decyzje, które następnie stają się częścią środowiska przyszłego uczenia. W długiej perspektywie pojawia się rekurencyjna pętla kulturowo-algorytmiczna, która nie musi prowadzić ani do postępu, ani do degeneracji.

Teoria ewolucji kulturowej uczy, że mechanizmy transmisji potrafią zachowywać wartościową wiedzę, ale mogą również wzmacniać błędy. Kluczowym pytaniem staje się zatem kwestia różnorodności źródeł, mechanizmów korekty, konkurencji hipotez i możliwości wprowadzania nowych informacji do systemu. Monokultura poznawcza może być wydajna, a jednocześnie krucha. Dlatego autonomia epistemiczna ma również wymiar zbiorowy. Społeczeństwo potrzebuje nie tylko inteligentnych jednostek, lecz instytucji pozwalających na kwestionowanie wyników, sprawdzanie źródeł i utrzymywanie pluralizmu modeli.

Sztuczna inteligencja, centralizująca coraz większą część pośrednictwa poznawczego, może zwiększać produktywność, lecz jednocześnie rodzi pytanie o systemową koncentrację epistemicznej infrastruktury. W tym punkcie teoria umysłu rozszerzonego spotyka się z teorią instytucji. Jeśli część praktycznego poznania zostaje delegowana technologii, kontrola nad nią staje się kontrolą nad fragmentem warunków poznawczego działania. Podobnie jak infrastruktura energetyczna jest warunkiem sprawstwa przemysłowego, tak infrastruktura informacyjna staje się warunkiem sprawstwa epistemicznego. Neurotechnologia dodatkowo zbliża tę infrastrukturę do biologicznego poziomu osoby. Można zatem mówić o powstawaniu warstwy egzosomatycznej podmiotowości: systemów znajdujących się fizycznie poza organizmem, lecz stale uczestniczących w procesach pamięci, komunikacji i podejmowania decyzji. Alfred Lotka używał pojęcia narzędzi egzosomatycznych w odniesieniu do technologii zwiększających możliwości organizmu poza jego biologiczne wyposażenie.

Współczesna kognitywistyka pozwala rozszerzyć tę intuicję na funkcje poznawcze. Człowiek nie tylko posługuje się sztucznymi narzędziami; coraz częściej organizuje własne życie tak, że bez nich jego funkcjonalna architektura wygląda zupełnie inaczej. Nie oznacza to narodzin „postczłowieka” w silnym sensie metafizycznym. Bardziej adekwatne jest stwierdzenie, że człowiek zawsze był gatunkiem współewoluującym z własnymi technologiami, lecz gęstość sprzężenia osiąga nowy poziom. Pismo rozszerzało pamięć, ale nie było wszczepione do mózgu. Smartfon organizuje dostęp do informacji, lecz nie dekoduje aktywności kory ruchowej. Adaptacyjna DBS i dwukierunkowe BCI przesuwają technologię wewnątrz pętli regulacyjnych, które wcześniej należały niemal całkowicie do domeny biologii.

## Konieczność rozróżnienia infrastruktury osobowej od urządzeń zewnętrznych

W tym kontekście szczególnego znaczenia nabiera kategoria **granicy funkcjonalnej osoby**. Choć anatomiczna granica skóry nie zawsze wyznacza kres procesu poznawczego, to z drugiej strony granica ta nie może zostać dowolnie rozszerzona na każdy system, z którym człowiek wchodzi w interakcję. Konieczne są zatem kryteria integracji: ciągłość, niezawodność, dwukierunkowe sprzężenie, wykorzystanie do kontroli zachowania oraz możliwość uznania danego procesu za własny. Im silniej występują te właściwości, tym bardziej uzasadnione staje się traktowanie technologii jako elementu funkcjonalnej architektury podmiotu.

Prowadzi to do paradoksu własności. Jeśli implant funkcjonalnie stanowi część systemu mojego sprawstwa, lecz prawnie pozostaje urządzeniem podległym reżimowi producenta, licencji i

oprogramowania, powstaje rozdzźwięk między **tożsamością funkcjonalną i własnością instytucjonalną**. Człowiek staje się zależny od komponentu, którego kodu nie kontroluje. Problem ten jest znany z technologii konsumenckiej, jednak w przypadku urządzeń regulujących mowę, ruch czy nastrój nabiera on jakościowo innego wymiaru. Prawo przyszłości będzie zatem musiało odróżnić zwykłe urządzenie od **infrastruktury osobowej**. Nie każda rzecz wszczepiona do ciała automatycznie staje się częścią osoby w sensie prawnym, ale nie wszystkie implanty można traktować jak zwykłą elektronikę użytkową. Im bardziej urządzenie warunkuje podstawowe funkcje psychiczne lub komunikacyjne, tym silniejszy jest argument za szczególną ochroną ciągłości jego działania, danych i dostępności. Choć rozważania te mogą sprawiać wrażenie futurystycznych, ich załączek odnajdujemy już w analizach autora.

Historia Davida pokazuje, że niewielka farmakologiczna ingerencja w system nagrody może wpływać nie tylko na wynik testu, ale na pracę, relacje, dbałość o siebie i społeczną pozycję człowieka. Przykład Wahida wskazuje zaś na DBS jako możliwą metodę modulacji określonych obwodów w trudnych przypadkach. Kierunek jest zatem widoczny: **terapia mózgu jest zarazem ingerencją w warunki społecznej biografii**, nawet jeśli celem klinicznym pozostaje konkretny objaw. Najskuteczniejszą ochroną przed błędnymi wnioskami jest tutaj zachowanie wielopoziomowej teorii osoby, wypracowanej w poprzednich częściach książki. Nie powinniśmy twierdzić, że lek „przywrócił duszę”, elektroda „zmieniła osobowość” czy AI „przejęła myślenie”, nie rozkładając tych stwierdzeń na bardziej precyzyjne mechanizmy. Należy pytać, która konkretnie funkcja uległa zmianie: wartościowanie nagrody, zdolność inicjowania, kontrola impulsów, pamięć robocza, komunikacja, metapoznanie, społeczne uczestnictwo czy narracyjna interpretacja siebie. To właśnie wielopoziomowość chroni zarówno przed redukcjonizmem, jak i przed techno-mistyką.

Ropinirol nie „dodaje ambicji” jako substancji psychicznej; zmienia warunki transmisji dopaminergicznej w systemie, którego reorganizacja może zwiększyć skłonność do działania. BCI nie „odczytuje duszy”, lecz rekonstruuje określone wzorce neuronalne związane z próbą mówienia. AI nie „posiada naszych myśli” tylko dlatego, że przewiduje tekst; może jednak realnie wpływać na proces tworzenia reprezentacji, jeśli zostanie głęboko włączona w poznawczą rutynę.

## **Suwerenność funkcjonalna jako warunek podmiotowości**

Z tej perspektywy najbardziej użyteczną kategorią dla przyszłości staje się **suwerenność funkcjonalna podmiotu**. Nie chodzi tu o suwerenność absolutną, gdyż człowiek zawsze pozostaje zależny od biologii, innych ludzi i instytucji. Chodzi raczej o zachowanie dostatecznej

zdolności do określania własnych celów, rozumienia warunków ich realizacji, kontrolowania technologii, odrzucania jej rekomendacji oraz możliwości odzyskania sprawstwa w razie wystąpienia błędu. Im bardziej technologia przenika system "Ja", tym silniejszy powinien być wymóg, aby jej architektura służyła użytkownikowi, a nie czyniła go funkcjonalnym dodatkiem do własnej logiki optymalizacyjnej. Przyszła neuroetyka powinna zatem stać się nie tyle etyką zakazującą ingerencji w mózg, co **teorią warunków podmiotowego korzystania z technologii umysłu**. Jej przedmiotem będą autonomia, autentyczność, ciągłość, mental privacy, sprawiedliwość, bezpieczeństwo, możliwość wycofania zgody, dostępność oraz rozkład odpowiedzialności. Właśnie wokół tych wartości – godności, integralności mentalnej, tożsamości, autonomii i prywatności – UNESCO koncentruje debatę nad zagrożeniami wynikającymi z rozwoju neurotechnologii połączonych z AI. Najgłębszy wniosek autora jest jednak bardziej antropologiczny niż regulacyjny.

Neurotechnologie nie dowodzą, że człowiek staje się "sztuczny" dopiero teraz. Ujawniają raczej fakt, który wcześniej łatwiej było ukryć: **podmiotowość zawsze była zależna od mechanizmów, których świadome "Ja" samo nie kontrolowało i których samo nie stworzyło**. Nie wybraliśmy przecież naszych genów, języka, wczesnego środowiska, architektury hipokampa ani reguł uczenia dopaminergicznego.

Technologie jedynie zwiększają możliwość świadomego wejścia w ten system i jego celowej modyfikacji. Paradoks polega na tym, że im większą kontrolę nad własnym mózgiem uzyskujemy, tym mniej oczywiste staje się pytanie o to, co właściwie oznacza "kontrolować siebie". Jeśli podmiot zmienia mechanizm, z którego wyrastają jego przyszłe preferencje, działa jednocześnie jako autor i materiał własnego projektu. To problem **samoodnoszącej się sprawczości**.

Architekt i budowla są tym samym organizmem. Nie istnieje punkt całkowicie zewnętrzny, z którego można by neutralnie zaprojektować przyszłe "Ja". Dzisiejsze wartości określają kierunek pożądanej zmiany, lecz sama zmiana może przekształcić wartości, według których później ocenimy wcześniejszą decyzję. Neuroenhancement ujawnia zatem problem temporalnej konstytucji autonomii: **które "Ja" ma prawo decydować za które przyszłe "Ja"?** Pytanie to przypomina klasyczne dylematy dotyczące testamentów, dyspozycji medycznych czy zobowiązań długoterminowych, jednak neurotechnologia nadaje mu znacznie większą wagę. Być może dlatego najbardziej dojrzałym kryterium nie jest niezmiennność, lecz **zachowanie możliwości dalszej rewizji siebie**. Autonomiczna technologia nie powinna zamykać podmiotu w jednym, raz wybranym stanie, chyba że istnieją niezwykle silne powody terapeutyczne. Powinna pozostawiać przestrzeń na uczenie się, zmianę zdania, dostrajanie i wycofanie. Autonomia nie oznacza bowiem osiągnięcia ostatecznej wersji siebie, lecz trwanie w procesie, w którym człowiek zachowuje

możliwie szeroką zdolność odpowiadania na nowe racje. W tym punkcie neuroetyka spotyka się z filozofią egzystencji.

Człowiek nie otrzymuje gotowej tożsamości, którą technologia mogłaby jedynie zachować lub zniszczyć. Tożsamość przez całe życie wyłania się z biologii, pamięci, relacji, przypadku, decyzji i reinterpretacji przeszłości. Technologia staje się kolejnym czynnikiem tej dynamiki, a jej wyjątkowość polega na tym, że może coraz skuteczniej oddziaływać **na mechanizmy, za pomocą których wszystkie pozostałe czynniki są oceniane.**

## **Suwerenność podmiotu jako kryterium dopuszczalności neurointerwencji**

Dlatego pytanie o to, czy po neurointerwencji człowiek pozostaje sobą, nie doczeka się jednej, uniwersalnej odpowiedzi biologicznej. Należy je rozbić na pytania bardziej precyzyjne: czy zachowano ciągłość pamięci? Czy wcześniejsze zobowiązania są wciąż rozpoznawane? Czy jednostka potrafi ocenić swój nowy stan i go zakwestionować? Czy jej zachowanie reaguje na racje? Czy relacje z innymi pozostają rozpoznawalne? Czy aktualne cele można włączyć do dotychczasowej biografii? Czy zmiana została przyjęta dobrowolnie i czy istnieje możliwość jej korekty?

Dopiero odpowiedzi na tę konstelację pytań pozwalają mówić o ciągłości podmiotowości. W tym świetle przyszłość nie prowadzi ku prostemu wyborowi między „człowiekiem naturalnym” a „człowiekiem ulepszonym”. Bardziej prawdopodobny jest świat różnych stopni technologicznego sprzężenia, w którym jedni będą korzystać z systemów przywracających utracone funkcje, inni z urządzeń kompensacyjnych, a jeszcze inni z rozszerzeń poznawczych, które nie są uznawane za medycznie konieczne. Zasadniczym problemem nie będzie zatem ontologiczna czystość człowieka, lecz utrzymanie warunków, dzięki którym technologicznie wspomagany podmiot nadal może być autorem własnej biografii.

Można w tym miejscu sformułować tezę zamykającą ten etap rozprawy: nie każda ingerencja w mechanizm „Ja” odbiera człowiekowi siebie – niektóre interwencje przywracają właśnie zdolność bycia sobą. Nie każda technologiczna pomoc zwiększa autonomię; część z nich może podnieść skuteczność działania kosztem refleksyjnej kontroli. Nie każda zmiana osobowości oznacza utratę tożsamości, podobnie jak nie każda ciągłość zachowania gwarantuje zachowanie sprawstwa. Podstawowym przedmiotem oceny nie powinien więc być sam fakt technicznej ingerencji, lecz sposób, w jaki zmienia ona wielopoziomą strukturę pamięci, wartości, kontroli, relacji i zdolności odpowiadania na racje.

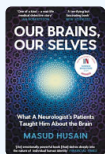
Wraz z tym wnioskiem analiza wchodzi na kolejny poziom. Początkowo pytaliśmy, jak drobne uszkodzenia mózgu mogą zmieniać człowieka. Następnie odkryliśmy, że „Ja” jest emergentną organizacją wielu systemów i relacji. Teraz okazuje się, że organizacja ta może być coraz częściej świadomie współprojektowana przez farmakologię, elektrostymulację, protezy neuronalne i sztuczną inteligencję. Najtrudniejsze pytanie przyszłości nie będzie zatem brzmiało, czy technologia stanie się częścią człowieka – w pewnym funkcjonalnym sensie już nią jest. Pytanie będzie brzmiało: jak zbudować takie sprzężenie człowieka z technologią, aby wzrost jego mocy nie prowadził do zaniku zdolności samodzielnego określania, czemu ta moc ma służyć?

To pytanie otwiera naturalnie kolejną część rozprawy. Po analizie indywidualnej autonomii należy przejść do społecznej i cywilizacyjnej skali modyfikowania człowieka: kwestii nierówności poznawczej, normalizacji ulepszania, komercjalizacji neurodanych, algorytmicznego kształtowania preferencji oraz nowych form zależności instytucjonalnej. Kluczowe będzie rozstrzygnięcie, czy prawo praw człowieka potrafi chronić podmiot w sytuacji, gdy przedmiotem ingerencji nie jest już wyłącznie ciało, własność albo informacja, lecz mechanizmy, za pomocą których człowiek tworzy własne przekonania, pragnienia i decyzje.

W świecie, gdzie granica między biologicznym „Ja” a algorytmicznym wsparciem ulega zatarciu, stajemy przed paradoksem samoodnoszącej się sprawczości: projektując swoje ulepszenia, zmieniamy jednocześnie architekta i materiał budowlany. Czy potrafimy stworzyć systemy, które zwiększając naszą moc, nie odbiorą nam prawa do bycia nieefektywnymi, omylnymi i w pełni autonomicznymi autorami własnych błędów? Ostatecznie największym wyzwaniem nie będzie pytanie o to, czy technologia stanie się częścią człowieka, lecz o to, czy człowiek w tym sprzężeniu pozostanie podmiotem, a nie tylko optymalizowanym dodatkiem do logiki systemu.

## Inspiracje

[1]



**"Our Brains Our Selves"** Masud Husain

ISBN: 9781837261109

**Masud Husain** jest brytyjskim neurologiem i neurobiologiem, profesorem neurologii i neuronauki poznawczej na Uniwersytecie Oksfordzkim oraz profesorem New College w Oksfordzie. Urodzony w Bangladeszu, wychował się w Wielkiej Brytanii. Ukończył studia medyczne i doktoranckie na Uniwersytecie Oksfordzkim, a następnie odbył staż podoktorski w MIT. Jego badania kliniczne i eksperymentalne koncentrują się na neurologii poznawczej, badając neurobiologiczne mechanizmy leżące u podstaw ludzkiej uwagi, pamięci roboczej, zaniedbywania przestrzennego i apatii w zaburzeniach takich jak udar, choroba Parkinsona i demencja. Husain jest wybranym członkiem Królewskiego Towarzystwa Naukowego i Akademii Nauk Medycznych oraz redaktorem naczelnym prestiżowego czasopisma Brain.

*Masud Husain is a British neurologist and neuroscientist serving as Professor of Neurology and Cognitive Neuroscience at the University of Oxford and a Professorial Fellow at New College, Oxford. Born in Bangladesh, he was raised in the United Kingdom and completed his medical degree and DPhil at the University of Oxford, followed by postdoctoral research at MIT. His clinical and experimental research focuses on cognitive neurology, investigating the neurobiological mechanisms underlying human attention, working memory, spatial neglect, and apathy in disorders such as stroke, Parkinson's disease, and dementia. Husain is an elected Fellow of the Royal Society and the Academy of Medical Sciences, and has served as the Editor-in-Chief of the prestigious journal Brain.*

*University of Oxford*

## Literatura uzupełniająca

### Książki

**Książki autora i dzieła przywołane w tekście:**



[1] **"Meet Mr. Product"**

Warren Dotz, Masud Husain

2003 | Chronicle Books | ISBN: 9780811835893



**[2] "Ad Boy"**

Warren Dotz, Masud Husain

2009 | ISBN: 9781580089845

**[3] "Alzheimer's Disease (AD) Detect & Prevent"**

Mia Dong, Masud Husain, David Brooks, Max Wilson, Michael Craven, Frédéric Destrebecq, Jean Georges, Kim Baden-Kristensen

2019

**[4] "Transport of Calcium and Magnesium Across the Placenta of the Diabetic Rat"**

Shahid Masud Husain, University of Manchester. Faculty of Medicine

1997

**[5] "Outsiderzy"**

Masud Husain

2026 | ISBN: 9788383678023

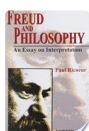
**Książki autorów przywoływanych i literatura pokrewna:**

**[6] "Claude Lévi-Strauss"**

Dan Sperber

**[7] "Oneself as Another"**

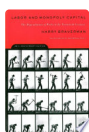
Paul Ricoeur



**[8] "Freud and Philosophy"**

Paul Ricoeur

Motilal Banarsidass Publishing House | ISBN: 9788120833050



**[9] "Labor and monopoly capital; the degradation of work in the twentieth century"**

Harry Braverman

NYU Press | ISBN: 9780853459408

**[10] "Elements of Physical Biology"**

Alfred Lotka

**[11] "Length of life; a study of the life table"**

Alfred Lotka

**[12] "The frequency distribution of scientific productivity"**

Alfred Lotka

## Artykuły naukowe

### Prace autorów przywoływanych w tekście:

**[1] "The moral, epistemic, and mindreading components of children's vigilance towards deception"**

Olivier Mascaró, Dan Sperber

2009 | Cognition | DOI: 10.1016/j.cognition.2009.05.012

[Źródło](#)

**[2] "Epistemic Vigilance"**

DAN SPERBER, FABRICE CLÉMENT, CHRISTOPHE HEINTZ, OLIVIER MASCARÓ, HUGO MERCIER

2010 | Mind & Language | DOI: 10.1111/j.1468-0017.2010.01394.x

[Źródło](#)

**[3] "Une pensée à l'orée des sciences cognitives"**

D. Sperber

2008

[Źródło](#)

**[4] "Time and narrative"**

Paul Ricœur, Kathleen McLaughlin, David Pellauer

1986

[Źródło](#)

**[5] "Hermeneutics and the Human Sciences"**

Paul Ricœur

2016 | Cambridge University Press eBooks | DOI: 10.1017/cbo9781316534984

[Źródło](#)

**[6] "Time and Narrative, Volume 1"**

Paul Ricœur, Kathleen McLaughlin, David Pellauer

1984 | University of Chicago Press eBooks | DOI: 10.7208/chicago/9780226713519.001.0001

[Źródło](#)

**[7] "Hermeneutics and the human sciences : essays on language, action and interpretation"**

Paul Ricœur

1981 | HAL

[Źródło](#)

**[8] "<i>Editorial Commentary</i> : Superior Capsular Reconstruction With Achilles Allograft Can Reverse Pseudoparalysis While Improving Functional Patient Outcomes"**

Alan W. Reynolds, Brian R. Waterman

2024 | Arthroscopy | DOI: 10.1016/j.arthro.2024.08.030

[Źródło](#)

**[9] "The Subject and Power"**

Michel Foucault

1982 | Critical Inquiry | DOI: 10.1086/448181

[Źródło](#)

**[10] "The Order of Things: An Archaeology of the Human Sciences."**

F. C. T. Moore, Michel Foucault

1971 | Man | DOI: 10.2307/2799252

[Źródło](#)

**[11] "Human Nature: Justice vs. Power (1971)"**

Noam Chomsky, Michel Foucault

2006 | The Chomsky-Foucault Debate | DOI: 10.2307/jj.32047651.4

[Źródło](#)

**[12] "The future as rupture: Michel Foucault"**

Michel Foucault

2011 | Time and Philosophy | DOI: 10.1017/upo9781844652778.013

[Źródło](#)

**[13] "The nature and value of rights"**

Joël Feinberg, Jan Narveson

1970 | The Journal of Value Inquiry | DOI: 10.1007/bf00137935

[Źródło](#)

**[14] "Rights, Justice, and the Bounds of Liberty"**

Joël Feinberg

1980 | Princeton University Press eBooks | DOI: 10.1515/9781400853977

[Źródło](#)

**[15] "The Expressive Function of Punishment"**

Joël Feinberg

1965 | The Monist | DOI: 10.5840/monist196549326

[Źródło](#)

**[16] "Doing and Deserving: Essays in the Theory of Responsibility."**

John L. Carafides, Joël Feinberg

1972 | Philosophy and Phenomenological Research | DOI: 10.2307/2106478

[Źródło](#)

**[17] "Analytical Note on Certain Rhythmic Relations in Organic Systems"**

Alfred J. Lotka

1920 | Proceedings of the National Academy of Sciences | DOI: 10.1073/pnas.6.7.410

[Źródło](#)

**[18] "UNDAMPED OSCILLATIONS DERIVED FROM THE LAW OF MASS ACTION."**

Alfred J. Lotka

1920 | Journal of the American Chemical Society | DOI: 10.1021/ja01453a010

[Źródło](#)

**[19] "RELATION BETWEEN BIRTH RATES AND DEATH RATES"**

Alfred J. Lotka

1907 | Science | DOI: 10.1126/science.26.653.21-a

[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: f9e1c820-dfb2-4a14-8817-d3c1ea879870