

Neurotechnologia a granice podmiotowości w świetle książki Our Brains Our Selves Masuda Husaina



AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje wpływ nowoczesnych technologii neurologicznych, takich jak DBS czy interfejsy mózg-komputer, na ludzką podmiotowość i poczucie tożsamości. Autor zastanawia się, gdzie przebiega granica między terapeutycznym przywracaniem sprawstwa a niepożądaną zmianą osobowości. Na przykładach farmakologicznych i technologicznych wykazano, że ingerencje w mózg mogą zarówno ograniczać autonomię, jak i paradoksalnie ją zwiększać, usuwając patologiczne bariery. Tekst podważa naiwne rozumienie autentyczności jako braku zewnętrznej ingerencji, sugerując, że kluczowa jest zgodność działań z wartościami pacjenta. Całość stanowi refleksję nad tym, jak techniczna rekonfiguracja mechanizmów neuronalnych zmienia pytanie o to, kim jesteśmy i co definiuje nasze „Ja” w dobie neurotechnologii.

The article analyzes the impact of modern neurological technologies, such as DBS or brain-computer interfaces, on human agency and the sense of identity. The author considers where the line is drawn between the therapeutic restoration of agency and an undesirable change in personality. Using pharmacological and technological examples, it is shown that brain interventions can both limit autonomy and, paradoxically, increase it by removing pathological barriers. The text challenges a naive understanding of authenticity as the absence of external interference, suggesting that alignment with the patient's values is key. Overall, it is a reflection on how the technical reconfiguration of neuronal mechanisms changes the question of who we are and what defines our 'Self' in the age of neurotechnology.

Artykuł analizuje fundamentalne przesunięcie w rozumieniu ludzkiej podmiotowości, wywołane przez rozwój neurotechnologii i sztucznej inteligencji. Autor stawia tezę, że

współczesne interwencje medyczne – od farmakologii po adaptacyjną głęboką stymulację mózgu (DBS) i interfejsy mózg-komputer (BCI) – nie są jedynie narzędziami naprawczymi, lecz instrumentami rekonfigurującymi strukturę ludzkiego sprawstwa. Kluczowym problemem przestaje być samo pytanie o skuteczność techniczną, a zaczyna być kwestia autentyczności i integralności psychicznej: czy zmiana zachowania wywołana technologicznie jest powrotem do „prawdziwego ja”, czy kreowaniem nowej, hybrydowej tożsamości. Tekst dowodzi, że w świecie zamkniętych pętli regulacyjnych i algorytmicznego wsparcia poznawczego, autonomia przestaje być stanem niezależności, a staje się procesem negocjowania kontroli nad systemami, które współtworzą nasze decyzje. W tym kontekście największym wyzwaniem etycznym nie jest lęk przed „utrata siebie”, lecz konieczność zdefiniowania nowej formy sprawstwa, w której człowiek zachowuje realny wpływ na funkcję celu sterującą jego własnym umysłem.

SŁOWA KLUCZOWE

neurotechnologia

podmiotowość

adaptive deep brain stimulation

interfejsy mózg-komputer BCI

neuroetyka

integralność psychiczna

prywatność mentalna

wolność poznawcza

cognitive offloading

rozproszone sprawstwo

funkcja celu

hybrydowa pętla sprawstwa

neuroprotezy mowy

architektura wyboru

Neurotechnologia jako narzędzie rekonstrukcji podmiotowości

Projektowanie "Ja": neurotechnologia, psychofarmakologia, sztuczna inteligencja i granice autonomii człowieka

Wgląd w chorobę prowadzi ku rekonstrukcji warunków podmiotowości. Uszkodzenie określonych obwodów ujawnia, jak wiele procesów musi współdziałać, aby człowiek pamiętał własną historię, rozpoznawał znaczenia, inicjował działanie, kontrolował impulsy, orientował się w przestrzeni i pozostawał uczestnikiem świata społecznego. Współczesna medycyna zmienia jednak kierunek tego eksperymentu. Neurobiologia nie tylko obserwuje rozpad funkcji, lecz coraz częściej potrafi ingerować w mechanizmy, z których funkcje te powstają. Farmakoterapia modyfikuje transmisję synaptyczną, głęboka stymulacja mózgu moduluje aktywność sieci, interfejsy mózg-komputer

tłumaczą aktywność neuronalną na mowę lub ruch urządzenia, a systemy sztucznej inteligencji przejmują część procesów pamiętania, wyszukiwania, planowania i formułowania wypowiedzi. Problem "kim jestem?" zyskuje więc nowy wymiar. Nie pytamy już wyłącznie, które mechanizmy tworzą "Ja", ale także co dzieje się z podmiotowością, gdy zaczynamy je technicznie rekonfigurować. Autor książki przygotowuje grunt do takiej analizy w sposób szczególnie interesujący.

Historia Davida pokazuje bowiem, że zmiana neurochemiczna może przekładać się na głęboką transformację zachowania. Po nieskuteczności lewodopy zastosowanie ropinirolu — agonisty receptorów dopaminowych D2/D3 — wiązało się z odzyskaniem inicjatywy, dbałości o siebie oraz aktywności zawodowej i społecznej. Ten sam przypadek przywołuje jednak przykład pacjenta, u którego działanie agonisty wywołało patologiczny hazard i utratę znacznych środków finansowych. Nie jest to jedynie farmakologiczna ciekawostka.

Odsłania ona centralny problem neuroetyki: jeżeli ten sam typ ingerencji może przywracać utraconą zdolność sprawstwa, ale w innych warunkach patologicznie przeskalować wartościowanie nagrody, to terapeutyczne znaczenie interwencji zależy nie tylko od tego, czy zwiększa ona aktywność określonego układu, lecz od tego, jak reorganizuje całą strukturę celów i kontroli osoby. To pierwszy powód, dla którego klasyczne przeciwstawienie "leczenie versus zmiana osoby" okazuje się zbyt proste. Każda skuteczna terapia zaburzenia psychicznego lub neurologicznego w pewnym sensie zmienia człowieka, skoro modyfikuje jego odczuwanie, zachowanie albo zdolności. Pacjent po skutecznym leczeniu depresji może inaczej oceniać przyszłość; osoba po opanowaniu padaczki odzyskuje możliwość prowadzenia aktywnego życia; człowiek z chorobą Parkinsona po stymulacji może poruszać się i funkcjonować w relacjach inaczej niż wcześniej. Gdyby każda zmiana psychologiczna naruszała tożsamość w znaczeniu normatywnie niepożądanym, skuteczna medycyna umysłu byłaby z definicji podejrzana właśnie dlatego, że działa. Dlatego współczesna neuroetyka odróżnia zmianę osobowości, tożsamości, autonomii oraz jakości życia. Są to kategorie częściowo nakładające się, lecz nie tożsame.

Najnowszy przegląd badań nad neuromodulacją wskazuje, że literatura dotycząca DBS intensywnie analizuje autonomię, świadomą zgodę, osobowość, autentyczność i tożsamość, traktując neurotechnologie nie tylko jako potencjalne zagrożenie dla autonomii, lecz również jako jej możliwe narzędzie przywracania. To kluczowe rozróżnienie: technologia ingerująca w mózg może ograniczyć autonomię, jeśli wytwarza zachowanie obce wartościom pacjenta, ale może ją zwiększyć, jeśli usuwa patologiczny mechanizm, który wcześniej uniemożliwiał realizowanie własnych celów. W przypadku Davida różnica ta jest widoczna niemal laboratoryjnie.

Autentyczność jako zgodność motywacji, a nie brak ingerencji

Farmakologia nie dostarczyła mu nowych kwalifikacji zawodowych ani nie nauczyła zasad życia społecznego. Autor opisuje raczej odzyskanie możliwości wykorzystywania kompetencji, które wcześniej pozostawały funkcjonalnie "odłączone" od spontanicznej inicjacji. Można zatem interpretować terapię jako odbudowę sprawstwa, jednak sama kategoria „powrotu do dawnego siebie” nie jest filozoficznie neutralna.

Zakłada ona bowiem, że stan wcześniejszy posiada szczególny autorytet normatywny. W wielu chorobach jest to intuicyjne: jeśli nagły udar zmienił sprawnego trzydziestolatka w osobę głęboko apatyczną, powrót do dawnych celów łatwo uznać za restaurację. Sytuacja komplikuje się przy schorzeniach wieloletnich, zaburzeniach rozpoczynających się wcześnie lub stanach, które pacjent zdążył już włączyć do własnego obrazu siebie. Nie zawsze istnieje jedno oczywiste „Ja sprzed choroby”, do którego medycyna miałaby obowiązek wrócić. Problem ten prowadzi do filozoficznego pojęcia autentyczności. W wersji naiwnej za autentyczne uznaje się to, co jest „naprawdę moje”, natomiast za nieautentyczne – to, co zostało „wprowadzone” przez lek, elektrodę lub algorytm. Takie kryterium zawodzi, gdyż wszystkie ludzkie preferencje mają swoje przyczyny: wychowanie, temperament, geny, doświadczenia społeczne, traumy, edukację, hormony czy neuroprzekąźnictwo. Farmakologicznie zmienione pragnienie nie jest ontologicznie mniej neuronalne niż pragnienie wcześniejsze. Problemem nie jest zatem sztuczność przyczyny, lecz relacja nowego stanu do refleksyjnych wartości i historii osoby oraz jej zdolności do uznania go za własny. W tym miejscu niezwykle użyteczna okazuje się teoria Harry'ego Frankfurta dotycząca pragnień pierwszego i drugiego rzędu.

Człowiek może czegoś pragnąć, a jednocześnie nie chce być osobą kierowaną przez to konkretne pragnienie. Uzależniony może odczuwać silny głód substancji, lecz na wyższym poziomie pragnąć uwolnić się od jego władzy. Odwrotnie, pacjent po leczeniu może stwierdzić: „wreszcie mogę chcieć tego, czego od dawna chciałam chcieć”. Autonomia nie jest wówczas prostą obecnością pragnienia, lecz częściową zgodnością pomiędzy warstwami motywacji. Neurointerwencja może tę zgodność zwiększać albo rozsadzać.

Nie wystarczy jednak zapytać pacjenta raz po zabiegu, czy „czuje się sobą”. Ocena autentyczności ma charakter temporalny. Człowiek może początkowo traktować nowy stan jako obcy, by po miesiącach włączyć go do swojej biografii; może też być odwrotnie – początkowa euforia może zostać później oceniona jako stan, w którym podejmowano decyzje niezgodne z długofalowymi wartościami. Badania osób korzystających z adaptacyjnej DBS pokazują, że empiryczna

rzeczywistość jest bardziej złożona niż sugeruje część spekulatywnej literatury neuroetycznej. W badaniu jakościowym uczestnicy w większości opisywali zmiany osobowości, nastroju i zachowania pozytywnie, a istotne trwale przemiany negatywne były rzadkie. Starszy przegląd dotyczący DBS w chorobie Parkinsona również nie znalazł mocnych podstaw do twierdzenia, że stymulacja typowo powoduje głęboką zmianę cech osobowości, choć podkreślał potrzebę lepszego badania adaptacji, celów i tożsamości narracyjnej. Neuroetyka powinna zatem unikać zarówno technofobicznej narracji o „elektrodzie przejmującej osobę”, jak i technologicznego optymizmu ignorującego realne ryzyka. Głęboka stymulacja mózgu jest pod tym względem szczególnym przypadkiem, ponieważ pozwala odróżnić ingerencję strukturalną od modulacji funkcjonalnej: elektroda nie zastępuje tkanki i zazwyczaj nie dostarcza gotowej treści psychicznej.

Adaptacyjna stymulacja mózgu tworzy hybrydową pętlę sprawstwa

Interwencja ta zmienia warunki aktywności określonych obwodów, przy czym efekt końcowy zależy od miejsca stymulacji, parametrów, stanu chorobowego, stosowanych leków oraz organizacji całej sieci. Autor książki wskazuje na DBS jako perspektywę modulowania dysfunkcyjnych obwodów w trudniejszych przypadkach neurologicznych. Współczesny rozwój technologii prowadzi jednak znacznie dalej: od stymulacji opartej na sztywno ustawionych parametrach ku systemom, które rejestrują sygnał neuronalny i dostosowują działanie do aktualnego stanu mózgu. W 2025 roku FDA zatwierdziła opcjonalną funkcję adaptive deep brain stimulation dla określonych systemów stosowanych w chorobie Parkinsona. System ten wykorzystuje biomarkery neuronalne, by modyfikować stymulację w sposób bardziej dynamiczny niż klasyczna, ciągła metoda DBS.

Z perspektywy inżynierii sterowania jest to przejście od modelu open-loop ku zamkniętej pętli regulacyjnej: urządzenie nie tylko oddziałuje na układ nerwowy, ale otrzymuje od niego informację zwrotną i na jej podstawie koryguje interwencję. Filozoficzne znaczenie tej zmiany wykracza poza techniczny opis. W klasycznej terapii pacjent przyjmuje lek lub korzysta ze stymulatora, a następnie funkcjonuje w zmienionym stanie. W układzie adaptacyjnym część regulacji zostaje przekazana systemowi, który w czasie rzeczywistym interpretuje sygnał biologiczny i decyduje o momencie oraz formie interwencji. Powstaje hybrydowa pętla sprawstwa: organizm wytwarza stan neuronalny, urządzenie go klasyfikuje i modyfikuje mózg, a ten z kolei generuje kolejne stany. Nie oznacza to, że urządzenie staje się osobą czy współautorem każdego działania pacjenta, lecz wskazuje na częściowe przesunięcie granicy między „terapią stosowaną przeze mnie” a „systemem

uczestniczącym w ciągłej regulacji mojego funkcjonowania”. Cybernetyka Norberta Wienera dostarcza tu zaskakująco aktualnego języka.

W systemie sprzężenia zwrotnego kontrola nie polega na jednorazowym wydaniu rozkazu, lecz na stałej obserwacji efektu i korekcie kolejnych działań. Organizm biologiczny funkcjonuje w ten sam sposób: reguluje temperaturę, ruch, ciśnienie, zachowanie i uwagę poprzez rozbudowane pętle sprzężenia. Adaptacyjna neurotechnologia nie wprowadza więc do człowieka całkowicie obcej logiki; wbudowuje sztuczny regulator w istniejący system regulacji biologicznej. Pytanie neuroetyczne brzmi zatem nie tylko: „czy urządzenie działa?”, ale przede wszystkim: „według jakiej funkcji celu działa i kto określił, jaki stan mózgu jest pożądany?”.

Pojęcie funkcji celu odsyła do teorii sterowania oraz uczenia maszynowego. Każdy regulator optymalizuje konkretny parametr: redukuje drżenie, stabilizuje biomarker lub zwiększa trafność dekodowania. Tymczasem życie człowieka składa się z wielu, często konkurencyjnych celów. Maksymalna redukcja jednego objawu może odbywać się kosztem energii, afektu, elastyczności poznawczej lub innych cech istotnych dla pacjenta. W medycynie jest to problem optymalizacji wielokryterialnej, w etyce zaś – konflikt wartości. Nie istnieje jedna skalarna funkcja „dobrego człowieczego życia”, którą można by wpisać do algorytmu. To właśnie dlatego medycyna oparta na wartościach pacjenta jest czymś więcej niż uprzejmym dodatkiem do technicznej optymalizacji.

Dwie osoby z podobnym objawem mogą zaakceptować odmienne kompromisy. Muzyk inaczej oceni subtelną zmianę w kontroli ruchu niż osoba wykonująca inny zawód; ktoś ceniący spontaniczność inaczej zinterpretuje skutki terapii niż osoba przedkładająca stabilność. „Najlepszy parametr kliniczny” nie musi być tożsamy z „najlepszym życiem dla danej osoby”. Problem ten koresponduje z capability approach Amartyi Sena i Marthy Nussbaum. Dobrostan w tym ujęciu nie polega wyłącznie na posiadaniu zasobu czy chwilowym odczuciu przyjemności, lecz na rzeczywistych możliwościach prowadzenia życia, które człowiek ma podstawy cenić.

Neurotechnologia może być analizowana jako narzędzie zwiększania tych możliwości: przywracanie komunikacji, ruchu, samodzielności czy udziału społecznego. Jednak samo zwiększenie jednej wydolności biologicznej nie gwarantuje automatycznie wzrostu całego zbioru możliwości. Technologia może bowiem tworzyć zależność od urządzenia, wymagać specjalistycznej opieki, generować koszty lub pozostawać niedostępna społecznie.

BCI a granica między dekodowaniem intencji a współtworzeniem wypowiedzi

Ta perspektywa ma kluczowe znaczenie dla interfejsów mózg-komputer. BCI nie są jedynie nowym sposobem „czytania w myślach”. Systemy implantowane, rozwijane w ramach praktyk klinicznych, dekodują konkretne wzorce aktywności neuronalnej i przekładają je na sterowanie kursorem, kończyną, urządzeniem lub generowanie komunikatu. Przegląd implantowanych BCI z 2025 roku, obejmujący dziesięciolecia badań nad przywracaniem czucia, kontroli ruchowej i komunikacji, podkreśla jednocześnie eksperymentalny charakter znacznej części tej dziedziny. Nie mamy tu do czynienia z telepatią, lecz z wysoce wymagającym statystycznym odwzorowaniem aktywności określonego obwodu na klasy sygnałów istotnych dla konkretnego zadania.

Najbardziej spektakularny postęp odnotowano w obszarze neuroprotez mowy. Przegląd w **Nature Reviews Neuroscience** opisuje systemy – rozwijane przede wszystkim dla osób z ciężkim paralizem – które dekodują aktywność związaną z próbą mówienia i przekładają ją na tekst, syntetyczny dźwięk lub animację twarzy. W 2025 roku opublikowano wyniki dotyczące interfejsu, który u mężczyzny z ALS przekładał aktywność z 256 mikroelektrod w korze ruchowej na syntetyzowaną mowę w czasie zbliżonym do rzeczywistego, umożliwiając modulowanie intonacji, a nawet prostych elementów śpiewu.

Z filozoficznego punktu widzenia znaczenie tego osiągnięcia jest fundamentalne: system nie nadaje człowiekowi „nowej osobowości”, lecz rekonstruuje kanał, dzięki któremu istniejąca intencja komunikacyjna może ponownie uzyskać społeczną skuteczność. Pojawia się jednak pytanie, gdzie kończy się dekodowanie zamiaru, a zaczyna współtworzenie wypowiedzi przez algorytm. Każda neuroproteza mowy musi bowiem mierzyć się z problemem niedoskonałego sygnału.

Współczesne systemy wykorzystują statystyczne modele języka, aby zwiększać trafność rekonstrukcji. Im bardziej aktywny staje się model, tym subtelniejsza staje się granica między tym, co zostało bezpośrednio przewidziane na podstawie wzorca neuronalnego, a tym, co algorytm uzupełnił w oparciu o prawdopodobieństwo językowe. Już we wspomnianym przeglądzie neuroprotez mowy modele językowe są definiowane jako jeden z elementów architektury technologicznej systemów dekodujących. Nie jest to jednak problem wyłącznie techniczny.

W codziennej komunikacji autorstwo wypowiedzi wiąże się z możliwością stwierdzenia: „to moje słowa”. Autokorekta, predykcja tekstu i asystenci językowi już dziś komplikują tę prostą deklarację. BCI czynią ten problem bardziej radykalnym, gdyż interfejs może działać jeszcze przed tradycyjną artykulacją. Jeśli dekodery błędnie wygeneruje zdanie, którego użytkownik nie zamierzał

wypowiedzieć, pojawia się kwestia semantycznej odpowiedzialności za wynik algorytmicznej rekonstrukcji intencji. Wówczas nie wystarczy zapytać o statystyczną dokładność systemu; należy ustalić, w jaki sposób użytkownik może zaakceptować, poprawić lub zawetować wygenerowany komunikat.

Z perspektywy filozofii działania kluczowy staje się zatem mechanizm kontroli korekcyjnej. Autorstwo nie wymaga być może samodzielnego wygenerowania każdego znaku – korzystając z języka naturalnego, i tak czerpiemy z zastanych struktur gramatycznych i słowników. Wymaga jednak realnej możliwości kontrolowania komunikatu na poziomie intencjonalnym: rozpoznania go jako zgodnego z zamiarem oraz możliwości interwencji w przypadku niezgodności. Przyszłe systemy neurotechnologiczne będą zatem oceniane nie tylko przez pryzmat **accuracy**, ale przez to, czy utrzymują meaningful human control nad rezultatem. Problem ten można ująć w kategoriach teorii rozproszonego sprawstwa: w złożonych systemach technicznych skutek jest bowiem generowany przez wiele elementów – człowieka, algorytm, dane treningowe, interfejs, sprzęt, operatora medycznego oraz instytucję. Nie oznacza to automatycznie całkowitego rozmycia odpowiedzialności.

Współzależne autorstwo i ochrona integralności psychicznej

W teorii organizacji i etyce technologii analizuje się raczej jej rozkład między uczestników, w zależności od ich zdolności do kontrolowania, przewidywania oraz korygowania działania systemu. BCI nie tworzy zatem „braku autora”, lecz strukturę współzależnego autorstwa, w której poszczególne role muszą zostać normatywnie odróżnione.

Jeszcze głębsze pytania pojawiają się w przypadku interfejsów dwukierunkowych. System jedynie odczytujący sygnały mózgu działa inaczej niż taki, który jednocześnie odczytuje i stymuluje. W pierwszym scenariuszu technologia przede wszystkim dekoduje; w drugim staje się częścią pętli przyczynowej, modyfikując stan, który następnie ponownie odczytuje. W 2025 roku czasopismo **Nature Medicine** opisywało rozwój dwukierunkowych BCI wspomaganych przez AI właśnie jako kierunek otwierający drogę do coraz intensywniejszej interakcji między algorytmem a procesami neuronalnymi. Na poziomie medycznym może to zwiększać skuteczność rehabilitacji lub przywracania funkcji, jednak na poziomie filozoficznym tworzy system, w którym granica pomiędzy regulacją endogenną i egzogenną staje się funkcjonalnie płynna.

Stąd wywodzi się pojęcie mental integrity – integralności psychicznej. UNESCO traktuje ją, obok godności, autonomii, prywatności mentalnej i tożsamości, jako jedną z centralnych kategorii etycznych neurotechnologii. Integralność nie oznacza tu absolutnej nietykalności mózgu – operacja neurologiczna przeprowadzona za świadomą zgodą może przecież służyć dobru pacjenta. Chodzi raczej o ochronę przed ingerencją nieautoryzowaną, manipulacyjną lub taką, która narusza zdolność do samodzielnego kształtowania własnych procesów mentalnych.

Z integralnością wiąże się mental privacy, czyli prywatność mentalna. Tradycyjne prawo do prywatności powstawało w świecie, w którym myśl niewypowiedziana pozostawała praktycznie niedostępna dla innych. Neurotechnologia osłabia tę techniczną gwarancję, choć należy unikać wizji z gatunku science fiction: współczesne systemy nie odczytują swobodnie pełnej treści umysłu. Potrafią jednak w określonych warunkach dekodować klasy zamiarów, ruchów, stanów sensorycznych czy prób mowy. W połączeniu z uczeniem maszynowym nawet częściowe neurodane mogą zyskać wartość inferencyjną znacznie większą niż ich surowa postać.

Helen Nissenbaum w teorii contextual integrity wykazała, że prywatność nie polega jedynie na kontroli nad tajemnicą, lecz na zgodności przepływu informacji z normami właściwymi dla danego kontekstu. Neurodane są tego doskonałym przykładem. Sygnał z implantu może być zasadnie udostępniony lekarzowi w celu dostrojenia terapii, ale ta sama informacja wykorzystana przez pracodawcę, ubezpieczyciela lub platformę reklamową narusza integralność kontekstu – nawet jeśli technicznie została „pozyskana legalnie”. Problem nie sprowadza się zatem do tego, czy dane są tajne, ale kto może z nich wyprowadzać jakie wnioski, w jakim celu i z jakimi konsekwencjami dla osoby. Jest to szczególnie istotne w synergii neurotechnologii z AI.

Algorytm nie musi otrzymać reprezentacji „myśli politycznej” czy „intencji zakupowej” jako gotowej kategorii. Wystarczy, że na podstawie danych biologicznych nauczy się przewidywać zachowanie z prawdopodobieństwem użytecznym ekonomicznie. Etyka danych definiuje ten problem jako prywatność inferencyjną: ryzyko nie wynika tu z ujawnienia informacji zapisanej wprost, lecz ze statystycznego wywnioskowania cechy, której podmiot nigdy świadomie nie udostępnił.

Kolejną kategorią jest cognitive liberty – wolność poznawcza. Jej sens jest dwoisty. Po pierwsze, człowiek powinien być chroniony przed niechcianą ingerencją w funkcjonowanie umysłu. Po drugie, pojawia się pytanie o zakres prawa do dobrowolnej modyfikacji własnego umysłu. Te dwie strony wolności mogą pozostawać w napięciu: nadmiernie paternalistyczny system może ograniczać autonomiczne korzystanie z technologii, natomiast skrajnie liberalny rynek może stworzyć środowisko, w którym formalnie dobrowolne ulepszanie staje się w rzeczywistości wymuszone przez konkurencję.

Granica między terapią a ulepszeniem w kontekście sprawiedliwości społecznej

To właśnie tutaj zaczyna się klasyczny problem terapia-enhancement. Czy interwencja jest terapią, gdy przywraca funkcję do statystycznej normy, a ulepszeniem, kiedy ją przekracza? Kryterium normy populacyjnej jest niewystarczające, ponieważ sama „normalność” ma charakter statystyczny, a nie moralny. Czy okulary poprawiające wzrok do wartości lepszej niż przeciętna stają się przez to moralnie innym przedmiotem? Czy lek poprawiający uwagę osoby z zaburzeniem jest terapią, podczas gdy ten sam preparat u osoby zdrowej uznajemy za enhancement, mimo że mechanizm biologiczny pozostaje identyczny? Granica ta zależy od celu, kontekstu klinicznego, ryzyka i społecznego znaczenia interwencji, a nie wyłącznie od chemii. Norman Daniels próbował wiązać terapię z przywracaniem normalnego funkcjonowania gatunkowego, argumentując, że choroba ogranicza sprawiedliwe szanse jednostki.

Zwolennicy bardziej liberalnych teorii enhancement, tacy jak Allen Buchanan czy Julian Savulescu, podkreślają natomiast, że nie każda poprawa przekraczająca medyczną normę jest moralnie podejrzana; jeśli bezpiecznie zwiększa ona zdolność człowieka do realizacji wartościowych celów, sama etykieta „ulepszenia” nie stanowi argumentu przeciwko niej. Krytycy, w tym Michael Sandel, zwracają uwagę na ryzyko przekształcenia życia w projekt kontroli oraz presję perfekcjonizmu. Spór ten nie ma prostego rozwiązania, lecz wszystkie strony trafnie wskazują, że problem nie sprowadza się do samej technologii, lecz dotyczy modelu dobrego życia i społecznych warunków wyboru. Ekonomia instytucjonalna i socjologia dostarczają tutaj argumentu, który neuroetyka czasem pomija.

Indywidualny wybór może bowiem zmieniać środowisko konkurencji. Jeśli pewna forma poznawczego ulepszania stanie się powszechna w danym zawodzie, rezygnacja z niej może oznaczać utratę szans. Dobrowolność formalna pozostanie, ale jej koszt wzrośnie. Jest to mechanizm podobny do problemów koordynacyjnych i wyścigu zbrojeń: decyzja racjonalna dla jednostki może prowadzić do systemu, którego żaden uczestnik pierwotnie nie preferował. Neuroenhancement nie jest zatem tylko pytaniem o prawo do ulepszenia, ale również o to, jak taka decyzja zmienia pole wyboru innych ludzi. John Rawls postawiłby tu pytanie o sprawiedliwość strukturalną, natomiast capability approach – o rzeczywisty rozkład możliwości.

Jeżeli zaawansowane technologie poznawcze będą dostępne tylko dla części społeczeństwa, mogą one nie tylko pogłębiać istniejące nierówności majątkowe, ale przekształcać je w nierówności zdolności poznawczych i sprawczości. UNESCO właśnie dlatego wymienia dostępność i sprawiedliwość społeczną wśród głównych problemów etycznych neurotechnologii. Dyskusja o

ulepszaniu człowieka bez teorii dystrybucji jest niepełna. Równie ważne jest rozróżnienie capacity enhancement od performance enhancement. Technologia może zwiększać trwałą zdolność człowieka albo jedynie poprawiać wykonanie zadania w czasie jej używania. Kalkulator zwiększa poprawność obliczeń, lecz niekoniecznie zdolność wykonywania rachunków bez niego. Nawigacja GPS podnosi skuteczność dotarcia do celu, ale nie musi rozwijać wewnętrznej orientacji przestrzennej. Zewnętrzny system pamięci może poprawiać realizację zadania, jednocześnie ograniczając trening pamięci biologicznej.

AI jako niewidzialny architekt ludzkich decyzji

To rozróżnienie stanie się kluczowe dla oceny AI. Psychologia poznawcza opisuje ten mechanizm za pomocą pojęcia **cognitive offloading**. Człowiek od dawna przenosi część obciążenia poznawczego na otoczenie: pisze listy, robi notatki, korzysta z zegarów, map, archiwów i kalendarzy. Nie jest to oznaka degeneracji, lecz naturalny sposób adaptacji. Przegląd w **Nature Reviews Psychology** z 2025 roku pokazuje, że offloading może znacząco poprawiać efektywność zadań pamięciowych, ale generuje też koszt: gdy dostęp do zewnętrznie zapisanej informacji zostanie niespodziewanie utracony, wyniki mogą być gorsze niż u osób polegających na pamięci wewnętrznej. Wydajność systemu rozszerzonego może więc rosnąć równocześnie z malejącą odpornością jego biologicznego komponentu na odłączenie narzędzia. To prowadzi bezpośrednio do koncepcji Clarka i Chalmersa.

Ich teoria rozszerzonego umysłu nie zakłada, że każdy telefon staje się automatycznie fragmentem psychiki. Wskazuje raczej na możliwość funkcjonalnego włączenia zewnętrznego zasobu w proces poznawczy – pod warunkiem, że jest on stabilnie dostępny, systematycznie wykorzystywany i odgrywa rolę porównywalną z procesem wewnętrznym. W epoce generatywnej AI hipoteza ta przestaje być wyłącznie filozoficznym eksperymentem myślowym. System może nie tylko przechowywać informacje, lecz je organizować, streszczać, podsuwać alternatywy i uczestniczyć w tworzeniu tekstu. Jednak generatywna AI różni się od notesu istotnym szczegółem: **nie jest pasywnym magazynem**.

Notes zwraca zapis, który wcześniej wprowadziliśmy. Model językowy generuje nową treść na podstawie statystycznych regularności i kontekstu. Tym samym delegowanie funkcji do AI nie jest zwykłym rozszerzeniem pamięci; może być rozszerzeniem inferencji, interpretacji i generowania wariantów działania. Granica pomiędzy narzędziem pamięciowym a partnerem poznawczym staje się przez to mniej wyraźna. W psychologii metapoznania rodzi to problem **monitorowania pochodzenia poznawczego**. Człowiek powinien wiedzieć nie tylko, jaka informacja została mu przedstawiona, ale także czy pochodzi ona z jego własnej pamięci, dokumentu źródłowego, osoby

trzeciej czy probabilistycznej syntezy modelu. Jest to technologiczna odmiana analizowanego wcześniej **source monitoring**. Jeżeli użytkownik zaczyna przypisywać sobie argumenty, sformułowania lub wspomnienia podpowiedziane przez system, zaburzeniu może ulec epistemiczna granica pomiędzy własnym wytworzeniem i zewnętrznym wsparciem. Nie musi to być jednak katastrofalne.

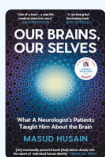
Ludzka kultura od zawsze jest kolektywna. Nikt nie konstruuje samodzielnie języka, pojęć naukowych ani większości argumentów filozoficznych, którymi się posługuje. Rozumowanie indywidualne zawsze wykorzystuje materiał przejęty od innych. Nowość AI leży raczej w **szybkości, personalizacji i niewidoczności pośrednictwa**. Książka nie dostosowuje każdej strony do czytelnika w czasie rzeczywistym. Algorytm może natomiast wybierać kolejność informacji, proponować sformułowania i kształtować zakres alternatyw, które użytkownik zobaczy. W tym miejscu teoria wyboru łączy się z pojęciem **choice architecture** Richarda Thaler'a i Cassa Sunsteina. Decyzja nigdy nie odbywa się w neutralnej próżni; sposób prezentacji opcji wpływa na wybór. Jeśli AI staje się stałym mediatorem poznawczym, może jednocześnie pełnić rolę architekta przestrzeni decyzyjnej.

Pytanie o autonomię nie dotyczy więc wyłącznie tego, czy człowiek może odrzucić rekomendację. Dotyczy również tego, **które możliwości w ogóle zostały mu przedstawione, jakie argumenty pominięto i według jakiej funkcji system określił ich kolejność**. Ekonomia behawioralna uczy, że ustawienie opcji domyślnej, rama zysku lub straty i sposób grupowania alternatyw mogą istotnie zmienić zachowanie bez jawnego przymusu. AI personalizujące informacje może te efekty zwielokrotnić, ponieważ nie musi stosować tej samej architektury wobec wszystkich osób. Z punktu widzenia autonomii pojawia się rozróżnienie pomiędzy **wsparciem decyzyjnym** a **sterowaniem decyzją**. Granica nie przebiega na poziomie tego, czy system "coś sugeruje", lecz na poziomie przejrzystości celu, możliwości kontroli, pluralizmu przedstawianych racji oraz podatności użytkownika na korektę własnych preferencji.

W świecie, gdzie granica między biologicznym impulsem a algorytmiczną korektą staje się płynna, musimy zapytać: czy dążąc do optymalizacji swojego funkcjonowania, nie zamieniamy wolności wyboru na komfort bycia doskonale sterowanym? Być może największym paradoksem nowoczesnej neurotechnologii jest to, że w pogoni za przywróceniem sprawstwa możemy stworzyć systemy, w których samo pojęcie 'autora' naszych decyzji stanie się jedynie wygodną iluzją. Czy jesteśmy gotowi zaakceptować wersję siebie, która jest efektem statystycznej optymalizacji, a nie organicznego błędu i walki?

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 nadzwyczajnym w New College w Oksfordzie. Jest również redaktorem naczelnym czasopisma Brain. Husain ukończył studia medyczne i doktoranckie na Uniwersytecie Oksfordzkim oraz staż podoktorski w Massachusetts Institute of Technology, zanim stworzył wiodący program badawczy w dziedzinie neurologii poznawczej. Jego badania kliniczne i naukowe dotyczą mechanizmów neuronalnych regulujących uwagę wzrokową, pamięć roboczą, podejmowanie decyzji i motywację. Jest szczególnie znany ze swoich przełomowych badań nad zaniedbaniem stronnym po udarze oraz neurobiologią apatii w chorobach neurodegeneracyjnych, w tym chorobie Alzheimera i Parkinsona. Husain jest wybranym członkiem Królewskiego Towarzystwa Naukowego i Akademii Nauk Medycznych.

Masud Husain is a British neurologist and neuroscientist serving as Professor of Neurology and Cognitive Neuroscience at the University of Oxford and Professorial Fellow at New College, Oxford. He also serves as the Editor-in-Chief of the journal Brain. Husain completed his medical and doctoral studies at the University of Oxford and postdoctoral training at the Massachusetts Institute of Technology before establishing a leading research program in cognitive neurology. His clinical and scientific research investigates the neural mechanisms governing visual attention, working memory, decision-making, and motivation. He is particularly renowned for his seminal investigations into hemispatial neglect following stroke and the neurobiology of apathy across neurodegenerative disorders, including Alzheimer's disease and Parkinson's disease. Husain is an elected Fellow of the Royal Society and the Academy of Medical Sciences.

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

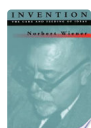
1997

[5] "Outsiderzy"

Masud Husain

2026 | ISBN: 9788383678023

Książki autorów przywoływanych i literatura pokrewna:



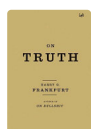
[6] "Invention: The Care and Feeding of Ideas"

Norbert Wiener

MIT Press | ISBN: 9780262731119

[7] "India: Economic Development and Social Opportunity"

Amartya Sen



[8] "On Truth"

Harry Frankfurt

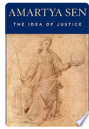
Random House | ISBN: 9781845951245



[9] "On Bullshit"

Harry Frankfurt

Princeton University Press | ISBN: 9780691276793



[10] "The Idea of Justice"

Amartya Sen

Harvard University Press | ISBN: 9780674036130

[11] "Obfuscation"

Helen Nissenbaum



[12] "Privacy in Context"

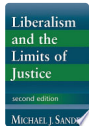
Helen Nissenbaum

2009 | Stanford Law Books | ISBN: 9780804752367

[13] "Contextual Integrity Through the Lens of Computer Science"

Sebastian Benthall, Seda Gurses, Helen Nissenbaum

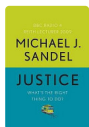
2017 | ISBN: 9781680833850



[14] "Liberalism and the Limits of Justice"

Michael Sandel

Cambridge University Press | ISBN: 9780521567411



[15] "Justice: What's the Right Thing to Do?"

Michael Sandel

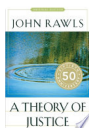
ISBN: 9781846142130



[16] "Justice as Fairness"

John Rawls

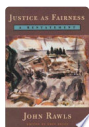
Belknap Press | ISBN: 9780674005112



[17] "A Theory of Justice"

John Rawls

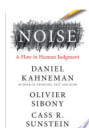
2005 | Harvard University Press | ISBN: 9780674017726



[18] "Justice as Fairness: A Restatement"

John Rawls

Harvard University Press | ISBN: 9780674005105



[19] "Noise: A Flaw in Human Judgment"

Cass Sunstein

Hachette UK | ISBN: 9780316451383



[20] "Too Much Information"

Cass Sunstein

MIT Press | ISBN: 9780262543910



[21] "Augmentation of Brain Function: Facts, Fiction and Controversy"

Manuel F. Casanova, Mikhail Lebedev, Ioan Opris

2018 | Frontiers Media SA | ISBN: 9782889456161



[22] "Psychological Abstracts"

Walter Samuel Hunter

1999

Artykuły naukowe

Prace autorów przywoływanych w tekście:

[1] "The Human Use of Human Beings: Cybernetics and Society"

Stephen B. Miles, Norbert Wiener

1951 | Land Economics | DOI: 10.2307/3159747

[Źródło](#)

[2] "The Human Use of Human Beings"

Norbert Wiener

1950

[Źródło](#)

[3] "IEEE 3rd Conference on Norbert Wiener in the 21st Century: Being Human in a Global Village"

N. Wiener, G. Adamson

2020 | IEEE technology & society magazine | DOI: 10.1109/mts.2020.3002051

[Źródło](#)

[4] "The challenge of gender justice"

Martha C. Nussbaum

2009 | Against Injustice | DOI: 10.1017/cb09780511657443.006

[Źródło](#)

[5] "Incoherent by Design? On the Moral Self-Consistency of LLMs"

Pegah Nokhiz, Aravinda Kanchana Ruwanpathirana, Helen Nissenbaum

2026 | arXiv (Cornell University) | DOI: 10.48550/arxiv.2608.15354

[Źródło](#)

[6] "Analyzing Privacy Policies Using Contextual Integrity Annotations"

Yan Shvartzshnaider, Noah Apthorpe, Nick Feamster, Helen F. Nissenbaum

2018 | DOI: 10.2139/ssrn.3244876

[Źródło](#)

[7] "Stop the Spread: A Contextual Integrity Perspective on the Appropriateness of COVID-19 Vaccination Certificates"

Shikun Zhang, Yan Shvartzshnaider, Yuanyuan Feng, H. Nissenbaum, N. Sadeh

2022 | Conference on Fairness, Accountability and Transparency | DOI: 10.1145/3531146.3533222

[Źródło](#)

[8] "Justice And Health care"

Allen Buchanan

2009 | DOI: 10.1093/oso/9780195394061.001.0001

[Źródło](#)

[9] "Justice as reciprocity versus subject-centered justice"

Allen Buchanan

1990 | DigitalGeorgetown (Georgetown University Library)

[Źródło](#)

[10] "Marx and Justice"

Allen Buchanan

1982

[Źródło](#)

[11] "A Critical Introduction to Rawls' Theory of Justice"

Allen Buchanan

2017 | Routledge eBooks | DOI: 10.4324/9781315257563-6

[Źródło](#)

[12] "THE MORAL OBLIGATION TO CREATE CHILDREN WITH THE BEST CHANCE OF THE BEST LIFE"

Julian Savulescu, Guy Kahane

2008 | Bioethics | DOI: 10.1111/j.1467-8519.2008.00687.x

[Źródło](#)

[13] "The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity"

Ingmar Persson, Julian Savulescu

2008 | Journal of Applied Philosophy | DOI: 10.1111/j.1468-5930.2008.00410.x

[Źródło](#)

[14] "Justice, Fairness, and Enhancement"

Julian Savulescu

2020 | The Ethics of Sports Technologies and Human Enhancement | DOI: 10.4324/9781003075004-26

[Źródło](#)

[15] "The importance of accurate representation of human brain organoid research"

M. Kataoka, C. Gyngell, Julian Savulescu, T. Sawai

2023 | Trends in Biotechnology | DOI: 10.1016/j.tibtech.2023.02.010

[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: 3e66fb11-1a4a-48f2-87ec-9247a7a870cf