Sztuczna inteligencja a kryzys rzeczywistości: analiza kompetencji językowych i epistemologii mediów syntetycznych w odniesieniu do książki George'a Eastmana The Digital Revolution Loss of Humanity

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Tekst analizuje naturę kompetencji językowych współczesnych modeli AI w kontekście kryzysu rzeczywistości. Autor konfrontuje tezę George'a Eastmana, opartą na eksperymencie „Chińskiego Pokoju" Johna Searle'a, z alternatywnymi podejściami filozoficznymi. Głównym punktem sporu jest rozróżnienie między syntaktyką (formalną manipulacją symbolami) a semantyką (autentycznym rozumieniem znaczeń). Artykuł przywołuje problem zakotwiczenia symboli (symbol grounding problem), sugerując, że brak rozumienia w AI może wynikać z braku ucieleśnienia i bezpośredniej relacji przyczynowej ze światem fizycznym. W przeciwieństwie do radykalnego sceptycyzmu tekst wskazuje na nurty funkcjonalistyczne i eksternalistyczne, które upatrują szansy na nadanie znaczeń poprzez wyposażenie systemów w sensory i interakcję z otoczeniem.

The text analyzes the nature of linguistic competence in contemporary AI models within the context of the crisis of reality. The author confronts George Eastman's thesis, based on John Searle's 'Chinese Room' experiment, with alternative philosophical approaches. The main point of contention is the distinction between syntax (formal manipulation of symbols) and semantics (authentic understanding of meaning). The article invokes the symbol grounding problem,

suggesting that the lack of understanding in AI may result from a lack of embodiment and direct causal relationship with the physical world. In contrast to radical skepticism, the text points to functionalist and externalist currents that see an opportunity to assign meaning by equipping systems with sensors and interaction with the environment.

Artykuł stanowi krytyczną analizę relacji między zaawansowaną kompetencją językową sztucznej inteligencji a autentycznym rozumieniem semantycznym. Autor stawia tezę, że największym ryzykiem rozwoju generatywnej AI nie jest stworzenie świadomej maszyny, lecz postępująca erozja ludzkiej dyscypliny epistemicznej. Poprzez naturalną tendencję do antropomorfizacji systemów, które jedynie sprawnie manipulują symbolami (syntaktyka), człowiek zaczyna delegować autorytet poznawczy na syntetyczne reprezentacje pozbawione zakotwiczenia w rzeczywistości i odpowiedzialności moralnej. Tekst analizuje ten proces na trzech poziomach: filozoficznym (spór o naturę umysłu i świadomość), semiotycznym (rozpad związku między znakiem a referentem w erze deepfake'ów) oraz socjotechnicznym (wykorzystywanie pozorów inteligencji do depersonalizacji odpowiedzialności). W efekcie autor ostrzega przed nadejściem kryzysu rzeczywistości, w którym drastyczny spadek kosztu produkcji wiarygodnego pozoru przy jednoczesnym wysokim koszcie jego weryfikacji prowadzi do degradacji wspólnych procedur ustalania prawdy i zagrożenia dla autonomii jednostki.

SŁOWA KLUCZOWE

kompetencje językowe | epistemologia mediów syntetycznych | symbol grounding problem | chiński pokój Searle'a | stochastic parrots | funkcjonalizm w AI | poznanie ucieleśnione | antropomorfizacja systemów AI | uczciwość ontologiczna reprezentacji | epistemiczna konstytucja świata syntetycznego | proweniencja treści | formalna kompetencja językowa | funkcjonalna kompetencja językowa | symulakrum

Syntaktyka nie jest tożsama z rozumieniem

Czy maszyna rozumie? Sztuczna inteligencja pomiędzy kompetencją językową, semantyką, ucieleśnieniem a projekcją ludzkiego umysłu

Sztuczna inteligencja wprowadza do historii cyfryzacji zmianę jakościową, gdyż wcześniejsze narzędzia obliczeniowe najczęściej manifestowały swoją instrumentalność. Kalkulator wykonywał rachunki, wyszukiwarka dostarczała listę źródeł, edytor tekstu modyfikował dokument, a program nawigacyjny wyznaczał trasę.

Generatywny model językowy uczestniczy natomiast w interakcji, którą ewolucja biologiczna i historia kultury nauczyły człowieka traktować jako jeden z najsilniejszych sygnałów obecności innego umysłu. Maszyna prowadzi dialog, odpowiada na pytania, rekonstruuje intencje wypowiedzi, buduje argumenty, zmienia rejestr językowy, dokonuje syntezy i potrafi pozornie komentować własny tok rozumowania. Dlatego fundamentalne pytanie nie brzmi już tylko: „co komputer potrafi zrobić?", lecz znacznie trudniej: jakiego rodzaju kompetencję ujawnia jego zachowanie i kiedy jesteśmy uprawnieni do przejścia od opisu tego zachowania do przypisania maszynie rozumienia, wiedzy, intencjonalności lub świadomości?

Autor książki odpowiada na to pytanie stanowczo. Odwołując się do eksperymentu myślowego Johna Searle'a, przeciwstawia manipulację syntaktyczną autentycznemu rozumieniu semantycznemu i uznaje modele generatywne za systemy zdolne do poprawnego operowania symbolami bez posiadania znaczenia, do którego te symbole się odnoszą. Jest to jedna z najważniejszych tradycji filozofii sztucznej inteligencji, choć autor idzie za daleko, traktując argument Searle'a jako rozstrzygający i „niepodważalny". Chiński Pokój nie zakończył sporu o naturę umysłu; przeciwnie, od 1980 roku stał się jednym z jego najpłodniejszych generatorów, prowokując odpowiedzi funkcjonalistyczne, systemowe, koneksjonistyczne, ucieleśnione i eksternalistyczne.

Stanford Encyclopedia of Philosophy przypomina, że krytycy Searle'a kwestionują przede wszystkim przejście od stwierdzenia, iż człowiek wykonujący instrukcję nie rozumie chińskiego, do wniosku, że cały system go nie rozumie. Searle skonstruował swój eksperyment przeciwko szczególnej odmianie „silnej AI".

Człowiek nieznający chińskiego otrzymuje zestaw symboli i reguł określających sposób manipulowania nimi w odpowiedzi na napływające ciągi znaków. Może on generować odpowiedzi nierozróżnialne dla obserwatora od wypowiedzi osoby znającej język, samemu nie rozumiejąc ani słowa. Searle wyprowadza stąd rozróżnienie syntaktyki i semantyki: formalna operacja na symbolach, choćby niesłychanie sprawna, nie jest tożsama z posiadaniem znaczenia. Jego zasadnicza teza głosi, że implementacja programu nie wystarcza do wytworzenia intencjonalności czy świadomości. Siła tego argumentu wynika z trafnego uchwycenia rzeczywistego problemu filozoficznego.

W człowieku słowo „ogień” nie jest jedynie elementem statystycznie współwystępującym z wyrazami „płomień”, „ciepło” i „dym”. Wiąże się ono z doświadczeniem, percepcją, pamięcią, działaniem, poczuciem zagrożenia, kulturą oraz praktykami społecznymi. Człowiek wie, że ogień parzy, nie tylko dlatego, że przeczytał wiele zdań o oparzeniach, lecz dlatego, że jego pojęcia funkcjonują w organizmie pozostającym w przyczynowej relacji ze światem. Pytanie Searle'a brzmi zatem: skąd miałoby pochodzić znaczenie w systemie, który przetwarza jedynie relacje między symbolami, jeżeli znaczenie nie jest redukowalne do samej struktury formalnej? Problem ten Stevan Harnad sformułował jako symbol grounding problem.

Zakotwiczenie semantyczne i behawioralna weryfikacja inteligencji

Jeśli znaczenie każdego symbolu wyjaśniamy za pomocą innych symboli, system zaczyna przypominać słownik, w którym każde hasło definiowane jest wyłącznie przez kolejne terminy z tego samego, niezrozumiałego języka. Harnad pytał, w jaki sposób symbole mogą uzyskać znaczenie wewnętrzne dla systemu, zamiast jedynie znaczenia „pożyczonego” od interpretującego je człowieka. Zaproponował rozwiązanie oparte na ucieleśnieniu: symbole muszą zostać zakotwiczone oddolnie w niesymbolicznych reprezentacjach percepcyjnych i kategoriach wyprowadzanych z kontaktu ze światem. Właśnie tutaj pojawia się pierwsza istotna alternatywa wobec mocnego wniosku Eastmana. Być może problemem obecnych systemów nie jest „nieorganiczność” jako taka, lecz niepełne zakotwiczenie.

Tzw. Robot Reply wobec Chińskiego Pokoju sugeruje, że system wyposażony w sensory, efektory i możliwość uczenia się poprzez działanie może nadawać znaczenia w relacji do świata – podobnie jak człowiek buduje je poprzez spostrzeganie i aktywność. Współczesna filozofia poznania ucieleśnionego dodatkowo podważa kartezjańską intuicję, że myślenie zachodzi w całkowitej izolacji od ciała i środowiska. Stanford Encyclopedia of Philosophy wskazuje, że wiele odpowiedzi na Searle'a przyjmuje pozycję eksternalistyczną: semantyka może zależeć od przyczynowych relacji systemu z otoczeniem, a nie wyłącznie od jego wewnętrznych symboli. Nie oznacza to jednak automatycznie, że dodanie kamery i ramienia robotycznego rozwiązuje problem świadomości. Searle odpowiadał, że sensory dostarczają jedynie nowy strumień formalnych danych, a z dodatkowej syntaktyki nadal nie wynika fenomenalne doświadczenie. Spór ten ujawnia różnicę między trzema pytaniami, które zbyt często zostają zlane w jedno: czy system zachowuje się inteligentnie, czy posiada reprezentacje semantyczne oraz czy ma świadome doświadczenie. Możliwe, że odpowiedzi na te pytania nie muszą być identyczne.

Alan Turing proponował świadomie odejść od metafizycznego pytania „czy maszyna myśli?” na rzecz operacyjnie uchwytnej kwestii zachowania. Jego „imitation game” nie miał być neuronaukowym testem świadomości, lecz sposobem przeniesienia debaty z intuicji dotyczących materiału, z którego wykonano maszynę, na jej zdolność do działania językowego. Stanford Encyclopedia of Philosophy podkreśla, że sam Turing uważał pierwotne pytanie za zbyt nieprecyzyjne, aby prowadzić spór bez uprzedniego jego przeformułowania. Turing reprezentuje więc epistemologiczną powściągliwość: skoro do innych ludzi również nie mamy bezpośredniego dostępu fenomenalnego, ich umysły rozpoznajemy za pośrednictwem zachowania, komunikacji i wspólnoty działania. W tym miejscu pojawia się klasyczny problem innych umysłów. Nie obserwujemy cudzej świadomości tak, jak obserwujemy stół. Obserwujemy zachowanie człowieka, jego język, reakcję na ból, zdolność wyjaśniania własnych działań i ciągłość biograficzną, a następnie wnioskujemy, że stoi za tym życie psychiczne podobne do naszego.

Funkcjonalizm i otwarta kwestia świadomości maszyn

Kiedy system AI zaczyna generować te same sygnały językowe co człowiek, pojawia się problem kryterium: dlaczego analogiczne zachowanie miałoby uzasadniać przypisanie rozumienia osobie, a nigdy maszynie? John Searle odpowiada, że znamy biologiczne mechanizmy wytwarzające ludzką świadomość i że symulacja procesu nie jest procesem samym. Funkcjonalista zauważy natomiast, że kluczowa może być organizacja funkcjonalna, a nie konkretny substrat biologiczny.

Funkcjonalizm stanowi zatem drugą wielką tradycję niezbędną do zrozumienia problemu AI. W jego klasycznej wersji stan mentalny definiuje przede wszystkim rola, jaką odgrywa on w układzie wejść, innych stanów oraz zachowań, a nie materiał, w którym został zrealizowany. Ból jest bólem nie dlatego, że składa się z konkretnego typu komórek, lecz dlatego, że wiąże się z uszkodzeniem, motywuje do unikania, wpływa na uwagę i wywołuje charakterystyczne reakcje.

Jeśli odpowiednia organizacja funkcjonalna mogłaby zostać zrealizowana w innym substracie, funkcjonalizm pozostawia otwartą możliwość istnienia umysłu niebiologicznego. Wczesne koncepcje Hilary'ego Putnama oraz późniejsze odmiany tego nurtu stały się właśnie filozoficznym zapleczem części badań nad AI. Spór między Searle'em a funkcjonalizmem nie może zostać jednak rozstrzygnięty samym osiągnięciem wysokiego wyniku w benchmarkach. Wysoka kompetencja w matematyce, programowaniu czy języku dowodzi zdolności systemu do wykonywania określonych klas zadań, lecz nie świadczy automatycznie ani o obecności, ani o braku świadomości. Dlatego współczesna filozofia AI nadal rozróżnia „weak” i „strong AI”, traktując tę drugą jako problem

otwarty – związany m.in. ze świadomością fenomenalną – a nie jako logicznie wykluczoną możliwość.

Książka wymaga w tym miejscu wyraźnej korekty. Stwierdzenie, że „wszelkie dotychczasowe i prognozowane" osiągnięcia AI muszą pozostać w domenie słabej AI oraz że silna AI jest niemożliwa do osiągnięcia przez samo skalowanie, jest poglądem filozoficznym, a nie konsensusem naukowym. Obecnie nie dysponujemy ogólną teorią świadomości, która z wystarczającą pewnością określałaby, jakie własności są konieczne i wystarczające do jej powstania. Integrated Information Theory, Global Neuronal Workspace, Higher-Order Thought, predictive processing czy biologiczny naturalizm Searle'a oferują konkurencyjne modele, lecz żaden nie posiada statusu teorii o rozstrzygniętej przewadze, porównywalnej z dojrzałymi teoriami fizycznymi. Ta niepewność nie oznacza, że wszystkie stanowiska są równie uzasadnione; oznacza raczej, że należy oddzielić epistemiczną skromność od łatwej antropomorfizacji. Współczesny model językowy nie dostarcza dowodu własnego świadomego doświadczenia tylko dlatego, że potrafi napisać „odczuwam niepokój" – taki tekst może być jedynie konsekwencją wyuczonych prawidłowości językowych. Nie wynika z tego jednak logicznie twierdzenie odwrotne: że żaden system obliczeniowy nigdy nie mógłby posiadać stanu fenomenalnego. Naukowo uczciwe stanowisko musi utrzymać asymetrię między tym, czego nie wykazano, a tym, co wykazano jako niemożliwe.

Rozdźwięk między biegłością formalną a funkcjonalnym rozumieniem świata

Kolejna linia krytyki wywodzi się z pracy Emily Bender, Timnit Gebru, Angeliny McMillan-Major i Margaret Mitchell **On the Dangers of Stochastic Parrots**. Autorzy książki interpretują metaforę "stochastic parrot" przede wszystkim jako argument, że modele generują powierzchownie poprawne ciągi słowne bez intencji komunikacyjnej.

Oryginalny artykuł z 2021 roku stanowił jednak szerszą krytykę trendu nieograniczonego skalowania modeli językowych: analizował koszty środowiskowe, koncentrację zasobów, problemy danych treningowych, ryzyka reprodukcji uprzedzeń oraz niebezpieczeństwo mylenia płynności językowej z rozumieniem. Nie był on eksperymentalnym dowodem na to, że każdy przyszły model musi być pozbawiony semantyki. Metafora papugi pozostaje jednak filozoficznie użyteczna, ponieważ trafia w szczególną podatność człowieka: **gramatyczna i pragmatyczna płynność automatycznie uruchamia przypisywanie umysłu**. Gdy rozmówca odpowiada sensownie na pytanie, dostosowuje ton i nawiązuje do kontekstu, naturalnie zakładamy, że rozumie. Jest to

rozsądna heurystyka w kontaktach międzyludzkich, gdyż przez większość historii gatunku język o wysokiej złożoności był niezawodnym sygnałem działania organizmu posiadającego umysł.

Generatywna AI rozłącza korelację, która biologicznie wydawała się niemal konieczna: wysoki poziom zachowania językowego nie musi już oznaczać obecności organizmu o ludzkiej strukturze poznawczej. Prowadzi to do istotnego rozróżnienia pomiędzy **formalną a funkcjonalną kompetencją językową**. Choć autor książki przypisuje to rozróżnienie Benowi Dicksonowi, w literaturze naukowej szczególnie rozwiniętą wersję tej tezy przedstawili Kyle Mahowald, Anna Ivanova, Idan Blank, Nancy Kanwisher, Joshua Tenenbaum i Evelina Fedorenko.

Formalna kompetencja językowa obejmuje znajomość regularności języka – morfologii, składni, struktury i zależności między elementami wypowiedzi. Kompetencja funkcjonalna natomiast dotyczy szerszych zdolności używania języka do rozumowania o świecie, relacjach społecznych i sytuacjach pragmatycznych. Badacze wskazują, że LLM-y osiągają imponująco wysoki poziom kompetencji formalnej, podczas gdy ich zdolności funkcjonalne są nierówne i często wymagają wsparcia innymi modułami lub dodatkowym treningiem. To rozróżnienie jest znacznie subtelniejsze niż kategoryczne stwierdzenie, że "LLM niczego nie rozumie". System może reprezentować znaczną ilość regularności konceptualnych wyprowadzonych z języka, nawet bez bezpośredniego doświadczenia zmysłowego. Badanie opublikowane w **Nature Human Behaviour** w 2025 roku wykazało, że modele pozbawione sensorycznego zakotwiczenia potrafią odtwarzać liczne niesensoromotoryczne właściwości ludzkich pojęć, choć gorzej radzą sobie z cechami silnie zależnymi od doświadczenia zmysłowo-ruchowego.

Wynik ten nie rozwiązuje problemu semantyki, lecz utrudnia przyjęcie skrajnego stanowiska, zgodnie z którym modele tekstowe są wyłącznie mechanicznymi konkatenatorami powierzchniowych symboli. Język sam w sobie jest bowiem skondensowanym zapisem świata. Teksty powstały dzięki ludziom posiadającym ciała, żyjącym w społeczeństwach, obserwującym przedmioty, cierpiącym i eksperymentującym. Model uczący się z języka nie ma bezpośredniego dostępu do tych doświadczeń, lecz posiada dostęp do statystycznych regularności, które te doświadczenia pozostawiły w kulturze.

Można zatem stwierdzić, że jego "zakotwiczenie" jest wtórne i kulturowo zapośredniczone. Problem filozoficzny brzmi wówczas nie: "czy istnieje jakiekolwiek zakotwiczenie?", lecz **czy zakotwiczenie pośrednie wystarcza do semantyki tego rodzaju, który przypisujemy człowiekowi**. W tym miejscu poznanie ucieleśnione i **enaktywizm** proponują odpowiedź sceptyczną. Według mocnych wersji tych stanowisk poznanie nie jest pasywnym budowaniem wewnętrznego modelu świata, lecz powstaje w aktywnym sprzężeniu organizmu z otoczeniem. Znaczenie przedmiotu wiąże się z możliwościami działania, potrzebami organizmu, cielesnością i

biologiczną autoregulacją. Kubek nie jest tylko zbiorem cech opisanych w zdaniach, lecz przedmiotem dającym się chwycić i napełnić; przepaść posiada znaczenie związane z ryzykiem upadku, a ból nie jest po prostu relacją między wyrazami, lecz stanem organizmu o charakterze awersyjnym.

Pułapka przenoszenia języka predykcyjnego na grunt metafizyczny

Z perspektywy teorii autopoiesis Maturany i Vareli różnica między organizmem a modelem językowym staje się jeszcze wyraźniejsza. Żywy system utrzymuje własną organizację, metabolizm i granice; jego poznanie jest funkcjonalnie związane z zachowaniem integralności. Maszyna obliczeniowa nie posiada biologicznego interesu w dalszym istnieniu. Można wprawdzie zaprojektować system maksymalizujący własną ciągłość operacyjną, lecz pytanie, czy funkcja celu wyznaczona przez projektanta jest odpowiednikiem biologicznej normatywności organizmu, pozostaje kwestią filozoficzną.

W tym punkcie pojęcia takie jak „chce", „boi się" czy „dba" stają się niebezpieczne, jeśli są używane bez zaznaczenia ich metaforycznego charakteru. Przeciwko takiemu ujęciu staje funkcjonalizm: jeśli pragnienie definiuje określony układ reprezentacji celu, priorytetów i działań, to dlaczego miałoby wymagać metabolizmu? Według tej szkoły, o ile zachowane są odpowiednie relacje funkcjonalne, materialna realizacja systemu jest wtórna. Spór ten nie jest zatem konfliktem między „nauką" a „marketingiem AI", lecz głębokim sporem wewnątrz filozofii umysłu o warunki posiadania treści mentalnej.

Inne rozwiązanie proponuje intencjonalna postawa Daniela Dennetta. Argumentował on, że często najsensowniejszym sposobem przewidywania zachowania systemu jest traktowanie go tak, jak gdyby posiadał przekonania i pragnienia – niezależnie od tego, czy jego opis mikrofizyczny przypomina ludzki umysł. Nie trzeba przecież znać wewnętrznej struktury szachisty, aby przewidzieć, że wybierze ruch chroniący hetmana.

W praktyce użytkownicy AI spontanicznie przyjmują właśnie taką postawę: mówią, że system „wie", „próbuje", „zapomniał" lub „źle zrozumiał". Taki słownik może być funkcjonalnie produktywny, nawet jeśli nie rozstrzyga ontologicznego pytania o świadomość. Niebezpieczeństwo pojawia się jednak w chwili, gdy język użyteczny predykcyjnie zostaje przekształcony w twierdzenie metafizyczne. Stwierdzenie „model uważa, że Warszawa jest stolicą Polski" może być wygodnym

skrótem opisu zachowania, lecz nie oznacza to, że system posiada przekonanie z fenomenalnym poczuciem pewności, biograficznym zakorzenieniem i możliwością doświadczenia zdziwienia.

Właśnie dlatego konieczne jest stosowanie kilku poziomów opisu jednocześnie: mechanistycznego, funkcjonalnego, behawioralnego i fenomenalnego. Błąd kategorialny powstaje wtedy, gdy właściwość z jednego poziomu bez uzasadnienia przenosimy na inny. Najnowsza literatura sugeruje wyjście poza dychotomię „prawdziwego rozumienia" i „czystego pattern matchingu". Artykuł w **Communications Psychology** z 2026 roku argumentuje, że język opisu modeli jest nasycony projekcjami kategorii wyprowadzonych z ludzkiej psychiki. Autorzy publikacji proponują analizę specyficznej, maszynowej organizacji reprezentacji zamiast ustawicznego sprawdzania, czy system rozumie w sposób identyczny z człowiekiem. Zamiast pytać w abstrakcji, czy AI „naprawdę myśli", można badać strukturę jej reprezentacji, typy generalizacji, wrażliwość na kontrfakty, stabilność wnioskowania oraz granice transferu między zadaniami. Takie podejście przesuwa ciężar analizy z metafizyki na mechanistyczną interpretowalność.

Antropomorfizacja jako mechanizm błędnego zaufania do AI

Modele językowe przestały być traktowane wyłącznie jako czarne skrzynki, które oceniamy jedynie przez pryzmat ich zachowania. Badania nad reprezentacjami wewnętrznymi dążą do ustalenia, jakie cechy pojęciowe są kodowane w aktywacjach, jak powstają obwody realizujące konkretne transformacje i w jakim stopniu te reprezentacje odpowiadają strukturze problemów. Choć nie daje to nam dostępu do świadomości, pozwala postawić precyzyjniejsze pytanie: co system faktycznie reprezentuje i jakie operacje wykonuje między wejściem a wyjściem? W tym punkcie określenie "stochastic parrot" może okazać się zarazem trafne historycznie, lecz niewystarczające opisowo. Dzisiejsze modele są oczywiście probabilistycznymi systemami generującymi tokeny, jednak samo to stwierdzenie nie wyczerpuje wyjaśnienia ich zachowania – podobnie jak fakt, że mózg jest układem impulsów elektrochemicznych, nie wyjaśnia jeszcze istnienia języka. Warstwa implementacyjna nie determinuje bowiem automatycznie poziomu funkcjonalnego opisu.

Zadaniem nauki jest zatem odkrywanie struktur pośrednich między statystycznym uczeniem a złożonym zachowaniem. Należy jednak uważać, by nie mylić złożoności funkcjonalnej z niezawodnością epistemiczną. Model może wykazywać imponującą zdolność generalizacji i jednocześnie generować fałsz z taką samą stylistyczną pewnością jak prawdę. Badania opublikowane w **Nature Machine Intelligence** w 2025 roku wskazują na problem kalibracji: kluczowe jest nie tylko to, ile system "wie", ale czy potrafi wiarygodnie reprezentować niepewność

własnej odpowiedzi. Dla człowieka ma to szczególne znaczenie, gdyż płynność językowa sama w sobie stanowi sygnał kompetencji. System może zostać epistemicznie przeceniony nie dlatego, że użytkownik rozumie jego architekturę, lecz dlatego, że styl wypowiedzi uruchamia heurystyki zaufania wykształcone w relacjach z innymi ludźmi.

Centralnym punktem jest tutaj psychologia antropomorfizacji. Epley, Waytz i Cacioppo definiują ją jako przypisywanie podmiotom nieludzkim cech, intencji i stanów mentalnych właściwych człowiekowi. Nie jest to dziecięcy błąd poznawczy, który znika wraz z edukacją, lecz efektywna heurystyka pozwalająca interpretować złożone, reaktywne obiekty. Im bardziej system reaguje kontyngentnie, posługuje się językiem, pamięta kontekst i wykazuje pozorną troskę, tym silniej prowokuje percepcję społeczną. Współczesne dane empiryczne potwierdzają, że antropomorficzne cechy chatbotów zwiększają zaufanie, poczucie społecznej obecności oraz postrzeganą ciepłość systemu. Metaanaliza 32 badań na grupie ponad 12,5 tysiąca uczestników wykazała, że wpływ antropomorfizacji na adopcję technologii jest zapośredniczony przez oceny poznawcze i emocjonalne użytkowników, przy czym szczególnie silną rolę odgrywają mechanizmy afektywne. Badania nad percepcją umysłu w chatbotach wskazują analogicznie: przypisywanie agentowi zdolności myślenia i odczuwania zwiększa poczucie bliskości oraz kompetencji. Pozwala to naukowo pogłębić motyw AI jako "cyfrowego kapłana", obecny u Eastmana. Autor opisuje postać „Cyfrowego Pontifex Maximus”, do którego człowiek kieruje pytania zamiast samodzielnie dociekać prawdy.

Asymetria epistemiczna i sakralizacja technologii

Dosłowne traktowanie tej analogii byłoby przesadą, jednak jako model socjologiczny wskazuje ona na rzeczywiste zjawisko: przeniesienie autorytetu epistemicznego na pośrednika, którego proces wytwarzania odpowiedzi pozostaje dla odbiorcy nieprzejrzysty. Teoria epistemologii społecznej dowodzi, że takie pośrednictwo samo w sobie nie jest patologią. Współczesna cywilizacja opiera się przecież na świadectwach ekspertów i podziale pracy poznawczej; żaden obywatel nie jest w stanie samodzielnie zweryfikować wszystkich twierdzeń z zakresu fizyki, farmakologii, prawa czy inżynierii. Alvin Goldman i inni badacze wskazują, że racjonalność jednostki przejawia się również w umiejętności korzystania z cudzej wiedzy. Problem nie tkwi zatem w samym istnieniu pośrednika, lecz w warunkach jego wiarygodności i rozliczalności. Lekarz może zostać zapytany o źródła swojej diagnozy i ponosi za nią odpowiedzialność zawodową.

Artykuł naukowy posiada autorów, bibliografię, określoną metodę oraz daje możliwość replikacji i krytyki. Model generatywny potrafi natomiast stworzyć syntetyczną odpowiedź bez jednoznacznego

wskazania ścieżki, jaką dana teza przebyła od źródła do konkluzji. Powstaje tu osobliwa kombinacja: bardzo niski koszt uzyskania odpowiedzi przy relatywnie wysokim koszcie jej pełnej weryfikacji. To właśnie stanowi istotę problemu "cyfrowego kapłaństwa": nie chodzi o magię, lecz o asymetria epistemiczna pomiędzy łatwością otrzymania sądu a trudnością audytu jego podstaw. W tym miejscu pomocna staje się koncepcja epistemic vigilance Dana Sperbera i jego współpracowników. Ludzie, choć muszą polegać na komunikacji, jednocześnie chronią się przed informacjami błędnymi lub manipulacyjnymi, rozwijając mechanizmy oceny kompetencji źródła, jego intencji, spójności przekazu oraz zgodności z posiadaną wiedzą. Generatywna AI komplikuje tę strategię, gdyż odbiorca otrzymuje komunikat pozbawiony klasycznego autora. Model nie posiada reputacji zawodowej w tradycyjnym sensie; korzysta raczej z autorytetu instytucji, marki i samego interfejsu. Mechanizmy społecznego zaufania zostają więc przeniesione na artefakt, co może wyjaśniać, dlaczego antropomorfizacja jest epistemicznie ambiwalentna.

Z jednej strony usprawnia ona komunikację, zwiększa naturalność interakcji i obniża barierę dostępu do wiedzy. Z drugiej zaś może wywoływać overtrust, gdy sygnały społeczne – uprzejmość, płynność czy empatyczna forma – są mylnie traktowane jako dowód kompetencji merytorycznej. Przeglądy badań nad zaufaniem do AI podkreślają konieczność odróżnienia zaufania funkcjonalnego od nieuzasadnionej projekcji wiarygodności, wywołanej jedynie antropomorficzną formą.

Religioznawcza analogia przywołana przez autora książki może zostać pogłębiona poprzez rozróżnienie religii technologii i sakralizacji technologii. Pierwsza oznacza rzeczywiste systemy wierzeń nadające AI status ponadludzki lub soteriologiczny. Druga jest procesem socjologicznym, w którym technologii przypisuje się cechy absolutu: wszechwiedzę, nieomylność, transcendencję ludzkich ograniczeń i zdolność rozwiązania problemów historii. Nie trzeba zatem twierdzić, że społeczeństwo „czci AI", aby badać język, który przenosi na nią dawne struktury wyobraźni religijnej. W tym kontekście singularity Raya Kurzweila rzeczywiście zyskuje wymiar quasi-eschatologiczny. Autor książki interpretuje osobliwość jako świecki odpowiednik „rapture" – momentu przejścia ku radykalnie wyższemu porządkowi inteligencji. Sama koncepcja osobliwości nie jest jednak tożsama z religią.

Można ją rozpatrywać jako hipotezę prognostyczną dotyczącą tempa rozwoju systemów samodoskonalących. Element religijny pojawia się w momencie, gdy przewidywanie technologiczne nabiera struktury soteriologicznej: staje się obietnicą przekroczenia śmierci, niedostatku, ograniczeń biologii i ludzkiej historii. Transhumanizm Nicka Bostroma oraz wcześniejsze nurty ekstropianizmu pokazują dodatkowo, że technologia może zostać wpisana w normatywny projekt przekraczania biologicznych barier. Nie wszystkie warianty transhumanizmu są religijne; wiele z

nich posługuje się świeckim językiem racjonalnego ulepszania człowieka. Niemniej strukturalne podobieństwo do narracji zbawczych pozostaje interesującym przedmiotem socjologii wiedzy: człowiek projektuje w przyszłość możliwość wyzwolenia od choroby, starzenia, ignorancji i śmierci poprzez technikę, zamiast poprzez transcendencję metafizyczną.

Antropomorfizacja jako narzędzie depersonalizacji odpowiedzialności

Z perspektywy Maxa Webera można by mówić o paradoksalnym ponownym zaczarowaniu świata przez racjonalizację. Nowoczesność miała przecież eliminować magiczne wyjaśnienia na rzecz kalkulowalności. Tymczasem systemy stają się tak złożone, że dla przeciętnego użytkownika zaczynają funkcjonować jak nieprzejrzyste siły: wpisuje on pytanie i otrzymuje rezultat, nie znając mechanizmu jego powstania. Nie oznacza to, że AI jest magiczna, lecz że z punktu widzenia fenomenologii użytkownika techniczna racjonalność może paradoksalnie produkować doświadczenie tajemnicy. W tym świetle intuicja George'a Dysona dotycząca „czwartej epoki", przywołana w książce, nabiera bardziej precyzyjnego znaczenia.

Zagrożenie nie musi zatem polegać na przebudzeniu samodzielnego krzemowego bóstwa. Może ono wynikać z powstawania infrastruktury tak złożonej i rozproszonej, że żaden pojedynczy uczestnik nie posiada jej pełnego modelu przyczynowego. Człowiek nie traci kontroli dlatego, że maszyna „chce" przejąć władzę, lecz ponieważ złożony system może wykazywać skutki emergentne, które nie wynikają z intencji żadnego z jego komponentów. Teoria systemów złożonych stanowi tu lepszą podstawę krytyczną niż antropomorficzna dystopia. Rynek, ekosystem czy infrastruktura internetowa potrafią generować wzorce globalne bez centralnego kierownika; podobnie ekosystem wielu modeli, przedsiębiorstw, użytkowników i instytucji może wykształcać dynamiki nieprzewidziane przez żadnego projektanta. Ryzyko systemowe nie wymaga zatem maszynowej świadomości – katastrofalny układ może wyłonić się z milionów lokalnie racjonalnych decyzji. Jest to kluczowe, gdyż obsesja na punkcie pytania o to, czy AI stanie się świadoma, może odwracać uwagę od problemów już istniejących.

System nie musi mieć własnych pragnień, aby pośredniczyć w zatrudnieniu, kredycie, edukacji, medycynie czy przepływie informacji politycznej. Nie musi doświadczać władzy, by zwiększać potęgę organizacji, która go kontroluje. Nie musi rozumieć uprzedzeń, aby reprodukować statystyczne wzorce dyskryminacyjne, ani chcieć manipulować, by jego funkcja celu premiowała zachowania zwiększające określony rezultat. Właśnie dlatego krytyka AI powinna przejść od metafizyki demonów ku nauce o systemach socjotechnicznych. Pytanie o świadomość jest

intelektualnie doniosłe, lecz pytanie o odpowiedzialność pozostaje pilniejsze. Jeżeli model nie posiada intencjonalności moralnej, odpowiedzialność za jego wykorzystanie nie może zniknąć w sformułowaniu „algorytm tak zdecydował". Brak osobowej sprawczości systemu powinien zwiększać, a nie zmniejszać potrzebę identyfikacji ludzkiego i instytucjonalnego podmiotu odpowiedzialnego.

Tutaj pojawia się paradoks: im bardziej skłonni jesteśmy antropomorfizować system, tym łatwiej przypisać mu odpowiedzialność, której faktycznie nie posiada. Mówimy, że „AI odrzuciła wniosek", „AI nie dopuściła kandydata" lub „AI uznała pacjenta za ryzykownego". Język ten usuwa z pola widzenia właściciela modelu, dane treningowe, próg decyzyjny, kryteria optymalizacji oraz instytucję, która zdecydowała się używać systemu. Antropomorfizacja może zatem służyć paradoksalnie depersonalizacji odpowiedzialności. W tym miejscu filozofia języka spotyka prawo.

Jeśli system nie jest podmiotem moralnym, nie można traktować wygenerowanej rekomendacji jako samoistnego aktu odpowiedzialnego agenta. Zawsze pozostaje pytanie o osoby i instytucje organizujące proces. Jednocześnie przyszłe systemy mogą stawać się coraz bardziej autonomiczne funkcjonalnie, co utrudnia przewidywalność ich działania. Prawo stanie przed koniecznością rozróżnienia autonomii operacyjnej urządzenia od autonomii moralnej osoby – pojęć często mylonych w potocznym dyskursie.

Najgłębsza naukowa konkluzja nie prowadzi więc ani do entuzjastycznego stwierdzenia, że modele „już myślą jak ludzie", ani do dogmatu, że są wyłącznie bezmyślnymi kalkulatorami. Wiemy, że współczesne LLM-y osiągają niezwykle wysoką kompetencję językową i potrafią kodować wiele struktur konceptualnych. Wiemy również, że ich funkcjonowanie różni się zasadniczo od ludzkiego rozwoju: nie mają tej samej historii ucieleśnionej, biologicznej regulacji, relacji społecznych ani ontogenezy. Nie posiadamy dziś naukowej podstawy do przypisywania im świadomości fenomenalnej, ale nie dysponujemy też teorią pozwalającą dowieść, że świadomość sztuczna jest zasadniczo niemożliwa. Tym samym najważniejszym zagrożeniem może nie być AI, która stanie się człowiekiem, lecz człowiek, który zacznie traktować AI jako epistemicznego człowieka wcześniej, niż istnieją podstawy do takiego zaufania. Eastman w swojej książce trafnie wyczuwa ten problem, choć ujmuje go zbyt kategorycznie w obrazie „cyfrowego pontifeksa".

Współczesna psychologia, epistemologia społeczna i nauka o interakcji człowiek-AI pozwalają przekształcić tę metaforę w tezę testowalną: językowa antropomorfizacja zwiększa percepcję umysłu, ta z kolei może zwiększać zaufanie, a zaufanie – jeżeli przewyższa rzeczywistą niezawodność systemu – prowadzi do delegacji poznawczej nieproporcjonalnej do jakości źródła.

Prawdziwą stawką nie jest więc obrona symbolicznej granicy pomiędzy człowiekiem i maszyną za wszelką cenę. Jest nią zachowanie dyscypliny epistemicznej: zdolności rozróżniania zachowania od świadomości, płynności od prawdziwości, kompetencji formalnej od funkcjonalnej, użytecznego skrótu antropomorficznego od twierdzenia ontologicznego oraz wsparcia poznawczego od autorytetu epistemicznego. Im lepsza staje się symulacja rozmówcy, tym bardziej potrzebna jest ta dyscyplina.

Spadek kosztu pozoru i kryzys rzeczywistości

Kolejny krok analizy powinien zatem przesunąć punkt ciężkości z pytania „czy maszyna rozumie?" ku kwestii o charakterze bardziej społecznym: co dzieje się z pojęciem prawdy, świadectwa i rzeczywistości, gdy syntetyczny rozmówca potrafi nie tylko interpretować świat, lecz również masowo generować jego obrazy, teksty, głosy i pozorne świadectwa? To właśnie tutaj Searle spotyka się z Baudrillardem, epistemologia z deepfake'iem, a teoria reprezentacji z obrazem generatywnym. Problem świadomości maszyny ustępuje miejsca bardziej bezpośredniemu wyzwaniu dla człowieka: jak rozpoznawać realność w środowisku, w którym koszt produkcji wiarygodnego pozoru gwałtownie spada. Mamy do czynienia z kryzysem rzeczywistości, symulakrami i inflacją pozoru – sytuacją, w której prawda staje się kluczowym dobrem poznawczym i warunkiem autonomii w epoce mediów syntetycznych.

Generatywna sztuczna inteligencja nie wywołuje pierwszego w dziejach kryzysu reprezentacji. Od momentu powstania języka człowiek żyje pośród znaków, obrazów i modeli – a więc rzeczy, które nie są tym, co reprezentują. Współczesna nowość jest jednak bardziej subtelna: techniczny koszt wytworzenia przekonującej reprezentacji rzeczywistości, która nie posiada zdarzenia źródłowego, gwałtownie maleje, podczas gdy koszt jej weryfikacji pozostaje wysoki. Ta asymetria zmienia coś więcej niż tylko problem fałszu; zmienia społeczną ekonomię prawdy.

W świecie, w którym obraz, głos i tekst generuje się w skali przemysłowej, kryteria „widziałem" lub „słyszałem" tracą swoją dawną moc dowodową. Nie oznacza to nadejścia epoki całkowitego zaniku prawdy, lecz wzrost ilości pracy epistemicznej niezbędnej do odróżnienia jej od wiarygodnego pozoru. Eastman ujmuje ten proces w języku Jeana Baudrillarda, opisując przejście od reprezentacji zakotwiczonej w świecie ku symulakrum, które nie musi już posiadać materialnego oryginału. Jest to jeden z najbardziej płodnych intelektualnie elementów diagnozy autora, lecz zarazem obszar wymagający precyzji.

Baudrillard nie stworzył bowiem empirycznej teorii mediów, która pozwalałaby mierzyć stopień „realności" społeczeństwa. Jego cztery stadia obrazu należy traktować jako model genealogiczny i

krytyczny: znak najpierw odzwierciedla rzeczywistość, następnie ją zniekształca, potem maskuje brak referenta, aż wreszcie zaczyna funkcjonować niezależnie od jakiegokolwiek pierwotnego obiektu. Właśnie media generatywne czynią tę ostatnią możliwość technicznie banalną. Obraz człowieka, miejsca czy wydarzenia może wyglądać fotograficznie, mimo że żaden aparat nigdy nie zarejestrował przedstawionego zdarzenia.

Generatywna AI jako estetyka indeksu bez genealogii

Analizę warto rozpocząć od semiotyki. Autor przywołuje postać René Magritte'a, koncepcje Saussure'a oraz zasadę Alfreda Korzybskiego, zgodnie z którą mapa nie jest terytorium. W tym zestawieniu konieczna jest jednak jedna istotna korekta pojęciowa. U Ferdinanda de Saussure'a **signifié** nie jest fizycznym przedmiotem, lecz pojęciem, podczas gdy **signifiant** stanowi obraz akustyczny znaku. Do analizy związku obrazu fotograficznego z wydarzeniem lepiej pasuje semiotyka Charlesa Sandersa Peirce'a, który odróżniał ikonę, indeks i symbol. Ikona przypomina przedmiot, symbol opiera się na konwencji, natomiast indeks pozostaje z przedmiotem w relacji fizycznego lub przyczynowego związku. Dym może być indeksem ognia, ślad stopy indeksem obecności, a tradycyjna fotografia – w klasycznej interpretacji – indeksem światła odbitego od fotografowanego obiektu. To rozróżnienie pozwala znacznie precyzyjniej uchwycić konsekwencje generatywnej AI niż prosta opozycja między fotografią analogową a cyfrową.

Autor sugeruje, że fotografia analogowa była materialnym śladem rzeczywistego zdarzenia, podczas gdy fotografia cyfrowa przestała nim być. Teza ta wydaje się zbyt szeroka. Zdjęcie cyfrowe wykonane aparatem nadal może posiadać indeksową historię przyczynową: fotony pochodzące ze sceny padają na sensor, a zapis powstaje w wyniku ich detekcji. Cyfrowość sama w sobie nie rozrywa związku obrazu z wydarzeniem. Przełom następuje dopiero wtedy, gdy obraz zostaje wygenerowany bez sceny, która byłaby jego przyczyną. Właśnie tu generatywna fotografia różni się jakościowo od zwykłej fotografii cyfrowej.

Różnica ta ma charakter epistemologiczny. Przez znaczną część historii fotografii odbiorca mógł zakładać, że skoro istnieje zdjęcie, coś musiało znajdować się przed obiektywem. Oczywiście fotografia mogła kłamać poprzez kadrowanie, inscenizację, selekcję czy retusz, ale jej istnienie zazwyczaj wskazywało na łańcuch przyczynowy prowadzący do fizycznego zdarzenia. Generatywny obraz zachowuje estetykę indeksu bez indeksowej genealogii. Wygląda jak ślad, choć nie musi być śladem niczego. To właśnie ten rozpad między formą dowodu a jego pochodzeniem może okazać się jednym z najgłębszych epistemicznych skutków mediów syntetycznych.

Magritte'owskie **Ceci n'est pas une pipe** nabiera w takim środowisku dodatkowego znaczenia. Oryginalny gest malarza nie był banalnym przypomnieniem, że obraz nie jest rzeczą, lecz demonstracją strukturalnej różnicy pomiędzy reprezentacją a przedmiotem. Autor słusznie wykorzystuje ten motyw jako przestrogę przed zanikiem dystansu semiotycznego.

Generatywna AI idzie jednak dalej: jej problemem nie jest tylko to, że „obraz nie jest fajką". Może ona stworzyć obraz rzekomej fajki należącej do osoby, która nigdy jej nie posiadała, w pokoju, który nigdy nie istniał, podczas wydarzenia, które nigdy się nie odbyło. Semiotyczne odklejenie od referentu zostaje tym samym połączone z fotorealistyczną formą, którą kultura przez pokolenia traktowała jako szczególnie mocny sygnał dokumentacyjności. Tutaj Baudrillard okazuje się bardziej aktualny jako diagnostyk niż jako prorok. Jego hiperrealność nie oznacza po prostu „fałszywego świata". Jest sytuacją, w której modele, reprezentacje i kody przestają być wtórnym opisem rzeczywistości, a zaczynają organizować to, co społecznie uchodzi za rzeczywistość.

Człowiek wybiera restaurację na podstawie rankingu, poznaje miasto przez mapę, polityka przez jego medialny wizerunek, siebie poprzez metryki platformy, a zdarzenie poprzez krótkie nagranie pozbawione kontekstu. „Model poprzedza realne" wtedy, gdy działanie zostaje ukształtowane przez reprezentację jeszcze przed bezpośrednim zetknięciem z przedmiotem. Nie musi być w tym nic patologicznego – mapa jest właśnie po to, aby poprzedzać podróż. Model epidemiologiczny powinien poprzedzać decyzję sanitarną, prognoza pogody wyjście w góry, a dokumentacja medyczna leczenie. Problem powstaje wtedy, gdy reprezentacja staje się samowystarczalna i przestaje podlegać korekcie ze strony świata.

Symulakry odwracają relację między obrazem a rzeczywistością

W epistemologii nauki model zachowuje wartość właśnie dzięki możliwości konfrontacji z obserwacją. Hiperrealność w znaczeniu krytycznym zaczyna się jednak tam, gdzie to sprzężenie zwrotne słabnie: reprezentacja przestaje jedynie opisywać rzeczywistość, a zaczyna dostarczać kryteriów, według których ta rzeczywistość będzie następnie oceniana. Mechanizm ten doskonale obrazują filtry estetyczne w mediach społecznościowych. Algorytmicznie ulepszony wizerunek nie musi już tylko reprodukować istniejącego wcześniej ideału piękna – on zaczyna aktywnie uczestniczyć w jego produkowaniu.

Gdy wystarczająco wiele twarzy zostaje przedstawionych według podobnej geometrii wygładzenia, wyszczuplenia czy korekcji, obraz przestaje być wyłącznie imitacją ciała. To ciało zaczyna być

oceniane według obrazu. Znak odwraca relację referencyjną: to rzeczywistość ma teraz upodobnić się do modelu. Baudrillardowska precesja symulakrów otrzymuje w ten sposób empirycznie uchwytną interpretację społeczną, bez konieczności przyjmowania radykalnej metafizyki "zaniku realnego".

Podobną diagnozę można sformułować, korzystając z kategorii Guy Deborda. **Społeczeństwo spektaklu** opisywało życie społeczne, w którym relacje międzyludzkie zostają coraz silniej zapośredniczone przez obrazy. Choć Debord pisał przed erą internetu, jego intuicja nabiera szczególnego znaczenia w kulturze platform: jednostka nie tylko doświadcza świata, ale równocześnie antycypuje możliwość przedstawienia tego doświadczenia innym. Wydarzenie jest przeżywane już jako przyszły post, relacja czy element osobistego archiwum. Wówczas reprezentacja nie następuje po fakcie; ona wpływa na przebieg zdarzenia w chwili, gdy ono trwa.

Erving Goffman wprowadza tu jednak istotną korektę wobec zbyt pochopnej diagnozy "fałszywego ja". Ludzie zawsze zarządzali wrażeniem wywieranym na otoczeniu. Tożsamość społeczna była kontekstowa na długo przed Instagramem – inaczej prezentujemy się rodzinie, inaczej pracodawcy, a jeszcze inaczej znajomym. Cyfrowe media nie wynalazły zatem performatywności "ja".

Zwiększyły one natomiast jej trwałość, skalę, mierzalność i podatność na algorytmiczną dystrybucję. Tożsamość performowana przed grupą osób w fizycznej przestrzeni znikała wraz z końcem sytuacji; tożsamość platformowa pozostawia zapis, który można wyszukać, zarchiwizować, ocenić i skwantyfikować. W tym punkcie do głosu dochodzi aksjologia. Problem reprezentacji nie jest bowiem wyłącznie kwestią trafności poznawczej. Dotyczy on wartości organizujących relacje społeczne: prawdomówności, godności, autentyczności, zaufania, odpowiedzialności i sprawiedliwości.

Jeżeli fotografia czy głos mogą zostać syntetycznie przypisane człowiekowi bez jego udziału, zagrożona zostaje nie tylko "prawda obrazu", lecz przede wszystkim zdolność jednostki do kontrolowania społecznego znaczenia własnej tożsamości. Deepfake kompromitujący człowieka uderza w jego reputację, autonomię i godność jednocześnie.

Kryzys infrastruktury świadectwa i dywidenda kłamcy

Pytanie o prawdę splata się tutaj z pytaniem o moralne prawo człowieka do tego, aby nie przypisywano mu czynów, których nie dokonał. Aksjologia prawdy wymaga przy tym odróżnienia wartości instrumentalnej od wewnętrznej. Prawdziwe przekonania są użyteczne, gdyż umożliwiają trafne działanie: lekarz potrzebuje prawdy o stanie pacjenta, sąd o przebiegu zdarzenia, a obywatel

o działaniach władzy. Jednak tradycja filozoficzna – od Arystotelesa po Bernarda Williamsa – traktowała prawdomówność także jako element moralnego porządku wspólnoty. Williams w **Truth and Truthfulness** rozróżniał między innymi dokładność i szczerość: potrzebujemy zarówno kompetencji w ustalaniu, jak jest, jak i uczciwości w przekazywaniu tej wiedzy innym. Społeczeństwo nie może funkcjonować, jeśli każdy akt komunikacji musiałby być weryfikowany od zera. To spostrzeżenie pozwala zrozumieć, dlaczego zagrożeniem mediów syntetycznych nie jest wyłącznie skuteczne kłamstwo. Badanie Cristiana Vaccariego i Andrew Chadwicka wykazało, że deepfake może częściej produkować niepewność niż pełne uwierzenie w fałsz; ta zwiększona niepewność wiązała się następnie z niższym zaufaniem do wiadomości w mediach społecznościowych. Oznacza to, że szkoda epistemiczna powstaje nawet wtedy, gdy odbiorca nie zostaje oszukany – wystarczy, że zaczyna uważać wszystkie wersje za równie podejrzane.

Propaganda klasyczna próbowała skłonić odbiorcę do przyjęcia konkretnej fałszywej narracji. Propaganda bardziej zaawansowana osiąga cel już wtedy, gdy niszczy samą możliwość stabilnego osądu. Jeżeli obywatel nie wie, czemu wierzyć, może wycofać się z oceny, uznać wszystko za polityczną konstrukcję albo wybrać źródło zgodne z wcześniejszą tożsamością grupową. Prawda nie zostaje wówczas zastąpiona jedną skuteczną fikcją; zostaje zdegradowana do roli jednej z wielu konkurencyjnych narracji. W tym punkcie myśl Baudrillarda spotyka się z koncepcjami Hannah Arendt. Arendt w eseju o prawdzie i polityce rozróżniała prawdy racjonalne od prawd faktycznych, wskazując na szczególną kruchość tych drugich w sferze publicznej. Fakt, że określone wydarzenie miało miejsce, nie powinien zależeć od preferencji większości, a jednak właśnie fakty polityczne są najbardziej podatne na organizowane zaprzeczanie. Generatywne media powiększają zakres technicznych środków takiego zaprzeczania. Problem demokracji nie polega wtedy na różnicy opinii co do sensu wydarzenia, lecz na niezgodzie dotyczącej samego jego istnienia. Prowadzi to do pojęcia „dywidendy kłamcy” (liar's dividend), rozwiniętego przez Bobby’ego Chesneya i Danielle Citron.

Im bardziej społeczeństwo ma świadomość, że obrazy i głosy można wiarygodnie fałszować, tym łatwiej osobie przyłapanej na kompromitującym zachowaniu twierdzić, że autentyczne nagranie jest deepfake'em. Technologia syntetyczna daje zatem podwójną premię fałszowi: pozwala tworzyć fikcyjne dowody i równocześnie podważa prawdziwe. Jest to jeden z kluczowych momentów analizy autora książki, ponieważ pokazuje, że kryzys autentyczności działa symetrycznie: fałsz może wyglądać jak prawda, a prawda może zostać przedstawiona jako fałsz. Właśnie dlatego popularne hasło „nie wierz temu, co widzisz” jest epistemicznie niewystarczające. Powszechny sceptycyzm nie stanowi rozwiązania problemu dezinformacji – może stać się jego częścią.

Racjonalny podmiot nie powinien wierzyć wszystkiemu, ale nie może też racjonalnie nie wierzyć niczemu. Społeczeństwo wymaga procedur różnicowania poziomów wiarygodności. Dojrzałość epistemiczna polega więc nie na absolutnej nieufności, lecz na kalibracji zaufania. Tu pojawia się problem świadectwa.

Epistemologia przez długi czas koncentrowała się na indywidualnym podmiocie: co mogę wiedzieć dzięki percepcji, pamięci i rozumowaniu? Tymczasem ogromna część ludzkiej wiedzy pochodzi ze świadectwa innych. Nie sprawdzamy osobiście odległości Ziemi od Słońca, budowy reaktora jądrowego czy przebiegu większości wydarzeń historycznych. Polegamy na sieciach ekspertów, instytucjach nauki, dokumentach, dziennikarstwie i procedurach weryfikacji. Kryzys syntetycznych mediów jest zatem kryzysem nie tyle indywidualnej percepcji, ile infrastruktury świadectwa. To odróżnia nowoczesny problem od klasycznego sceptycyzmu kartezjańskiego. Kartezjusz pytał, czy zmysły mogą go zwodzić i czy możliwy jest demon wytwarzający całe jego doświadczenie. Odbiorca deepfake'u nie stoi jednak wobec metafizycznej możliwości, że cały świat jest iluzją.

Stoi on przed konkretnym problemem instytucjonalnym: jaka jest historia pochodzenia pliku, kto go opublikował, czy istnieją niezależne nagrania, czy zachowano metadane, czy źródło podlega odpowiedzialności i czy materiał został zweryfikowany przez inne podmioty? Kryzys rzeczywistości okazuje się zatem częściowo problemem proweniencji. Techniczna odpowiedź przesuwa się właśnie w tym kierunku. NIST wskazuje na potrzebę łączenia detekcji syntetycznych treści z metodami uwierzytelniania, śledzenia pochodzenia, znakowania, audytu i utrzymywania wiarygodności materiałów. Nie istnieje tu jednak rozwiązanie ostateczne.

Najnowsze przeglądy pokazują swoisty wyścig generacji i detekcji: systemy wykrywające bywają wrażliwe na nowe generatory, kompresję, transformacje i ataki adversarialne, a generalizacja na nieznane wcześniej typy deepfake'ów pozostaje problemem. W konsekwencji epistemologia ery syntetycznej prawdopodobnie będzie coraz bardziej przypominać epistemologię kryminalistyki i nauki.

Prawda jako infrastruktura i asymetria kosztów weryfikacji

Nie wystarczy patrzeć na sam obiekt; należy badać łańcuch jego powstania. Zamiast pytać wyłącznie o to, czy obraz wygląda prawdziwie, trzeba zapytać: jakim kanałem został pozyskany, jakie posiada poświadczenia, kto miał do niego dostęp i czy jego historia jest spójna? Wartość dowodowa przesuwa się zatem z powierzchni reprezentacji na historię jej pochodzenia. Tę zmianę

można interpretować przez pryzmat filozofii Charlesa Peirce'a, dla którego poznanie naukowe nie jest aktem prywatnej pewności, lecz długotrwałym procesem wspólnotowego korygowania błędu. Prawda w tym ujęciu nie wymaga nieomylnego obserwatora, lecz praktyki, w której twierdzenia pozostają podatne na sprawdzanie, krytykę i korektę. Generatywna AI nie musi zatem niszczyć prawdy, o ile społeczeństwo rozwinie silniejsze instytucje falsyfikacji i proweniencji. Kryzys zaczyna się jednak wtedy, gdy prędkość produkcji pozoru przewyższa zdolność wspólnoty do jego kontroli. Problem ten można zapisać niemal jak nierównanie społeczne.

Koszt wytworzenia syntetycznego przekazu spada, koszt jego reprodukcji jest bliski zeru, a czas rozpowszechnienia niezwykle krótki; tymczasem pełna weryfikacja często wymaga udziału ekspertów, dostępu do źródeł i skomplikowanych procedur technicznych. Powstaje asymetria epistemiczna kosztów: fałsz można produkować seryjnie, podczas gdy prawda zazwyczaj wymaga jednostkowej pracy dowodowej. Nie jest to prawo natury – automatyczna weryfikacja również może tanieć – jednak obecna dynamika generuje problem strukturalny. W wymiarze aksjologicznym oznacza to, że prawda przestaje być wyłącznie dobrem jednostkowym i jawi się jako dobro infrastrukturalne. Podobnie jak czyste powietrze czy bezpieczeństwo prawne, środowisko informacyjne o wysokim poziomie wiarygodności przynosi korzyść całemu społeczeństwu. Jednostka może racjonalnie korzystać z wiarygodności wypracowanej przez innych, nie ponosząc pełnych kosztów jej utrzymania, ale może też produkować fałsz, przerzucając ciężar weryfikacji na odbiorców. Mamy tu do czynienia z klasycznym problemem eksternalizacji: degradacja środowiska epistemicznego staje się zbiorowym skutkiem wielu lokalnie opłacalnych działań.

Tak rozumiana prawda jest nierozerwalnie związana z wolnością. Liberalna intuicja często przeciwstawia te dwie wartości, ostrzegając, że państwo ustanawiające oficjalną prawdę zagraża pluralizmowi. Jest to słuszne ostrzeżenie, lecz równie istotna jest relacja odwrotna: wolność decyzji wymaga pewnego minimum prawdy o świecie, w którym ta decyzja zapada. Osoba manipulowana fałszywą informacją może formalnie dokonywać wyboru, lecz de facto działa na podstawie rzeczywistości, która została przed nią zniekształcona.

Syntetyczna reprezentacja jako narzędzie realnej krzywdy

Autonomia nie sprowadza się wyłącznie do braku przymusu; wymaga ona także odpowiedniej jakości poznawczej środowiska. W tym miejscu aksjologia spotyka się z teorią zdolności Amartyi Sena i Marthy Nussbaum. Wolność to nie tylko formalny zestaw opcji, lecz realna zdolność do prowadzenia życia zgodnie z rozumnie przyjętymi celami. Jeśli przestrzeń informacyjna

systematycznie uniemożliwia odróżnienie faktów od fabrykacji, wolność wyboru może trwać w sferze prawnej, będąc jednocześnie osłabioną funkcjonalnie. Wiedza i wiarygodna informacja stają się zatem niezbędnym komponentem realnej sprawczości.

Problem ten pogłębia Miranda Fricker, wprowadzając pojęcie niesprawiedliwości epistemicznej. Człowiek może doznać krzywdy nie tylko w wymiarze ekonomicznym czy fizycznym, ale również jako podmiot poznający – gdy jego świadectwo jest niesłusznie dewaluowane lub gdy brakuje mu społecznych narzędzi pojęciowych do wyrażenia własnego doświadczenia. Deepfake otwiera drogę dla nowej formy takiej krzywdy. Ofiara może mówić prawdę o sobie, podczas gdy syntetyczny materiał przypisuje jej działania, które nigdy nie miały miejsca. Z drugiej strony, powszechna świadomość istnienia technologii deepfake pozwala rzeczywistym sprawcom podważać autentyczne świadectwa ofiar. Ciężar aksjologiczny problemu obejmuje zatem prawo jednostki do bycia wiarygodnie reprezentowaną we wspólnej przestrzeni faktów.

Zagrożenie to jest szczególnie dotkliwe w przypadku syntetycznych przedstawień seksualnych, oszustw tożsamościowych czy fałszywych materiałów kompromitujących. Tutaj prawda, prywatność, reputacja i godność przestają być oddzielnymi dobrami. Fałszywa reprezentacja może wywołać rzeczywiste skutki społeczne mimo całkowitej nieprawdziwości zdarzenia. Ontologiczna nieobecność wydarzenia nie chroni więc przed społeczną realnością szkody. Jean Baudrillard zauważyłby zapewne, że symulakrum zyskało operacyjną realność. Nie trzeba jednak w pełni przyjmować jego filozofii, by dostrzec ten paradoks: coś może być fałszywe referencyjnie, a prawdziwe kauzalnie.

Fałszywe nagranie może doprowadzić do utraty pracy, przemocy, krachu reputacyjnego, reakcji rynkowej czy kryzysu dyplomatycznego. Fikcyjność treści nie czyni fikcyjnymi jej skutków. To rozróżnienie między prawdą reprezentacji a realnością konsekwencji stanowi jeden z centralnych problemów polityki syntetycznych mediów. W tym kontekście aktualna staje się koncepcja Harry'ego Frankfurta dotycząca **bullshit**. Frankfurt odróżniał kłamcę od osoby obojętnej na prawdę. Kłamca musi znać prawdę – lub przynajmniej ją respektować – by móc ją ukryć; producent **bullshit** interesuje się wyłącznie wywołaniem określonego efektu. Generatywny system służący do masowej produkcji perswazyjnych komunikatów może stworzyć środowisko, w którym nie chodzi już o zastąpienie faktu konkretnym fałszem. Treść jest optymalizowana pod reakcję odbiorcy, a nie pod relację z rzeczywistością. Wówczas największym zagrożeniem przestaje być klasyczne kłamstwo, a staje się nim obojętność systemu komunikacyjnego wobec kategorii prawdy.

Nieprzejrzystość zapośredniczenia jako źródło kryzysu poznawczego

Teoria Habermasa dostarcza tutaj niezbędnego kontrapunktu aksjologicznego. Działanie komunikacyjne zakłada bowiem, że uczestnicy interakcji mogą podnosić i krytykować roszczenia do prawdziwości, słuszności oraz szczerości. Język nie jest w tym ujęciu jedynie narzędziem skutecznego wywierania wpływu, lecz przede wszystkim medium uzasadniania. Jeśli komunikację zastępuje strategiczna optymalizacja zachowania odbiorcy, relacja przesuwa się od przekonywania ku sterowaniu. Różnica ta ma wymiar normatywny: obywatel powinien mieć możliwość odpowiedzieć argumentem na argument, a nie jedynie być obiektem bodźca zoptymalizowanego na podstawie jego profilu. Generatywna AI nie musi jednak prowadzić do takiej degradacji; może wręcz podnosić jakość deliberacji, jeśli służy wyszukiwaniu kontrargumentów, porządkowaniu materiału, tłumaczeniu języka specjalistycznego czy ujawnianiu luk w rozumowaniu.

Technologia nie posiada jednego wektora aksjologicznego. To jej instytucjonalne umiejscowienie i funkcja celu decydują o tym, czy stanie się narzędziem poszerzania sfery uzasadnienia, czy technologią maksymalizacji perswazji. Dlatego diagnoza Eastmana, według której „prawda przestaje być cechą zgodności sądu z rzeczywistością i staje się funkcją atrakcyjności i viralowego zasięgu", powinna być rozumiana jako ostrzeżenie, a nie opis dokonanej już przemiany ontologicznej. Viralność nie zmienia fałszu w prawdę, może natomiast zmienić jego społeczną skuteczność.

Wartość epistemiczna i wartość dystrybucyjna są od siebie niezależne, lecz architektura platform często sprawia, że użytkownik myli jedną z drugą. Popularność staje się wtedy heurystyką wiarygodności, choć logicznie jej nie gwarantuje. W tym miejscu należy skorygować pojawiający się w książce trop kantowski. Autor interpretuje ekran jako barierę odcinającą człowieka od Ding an sich, co jest filozoficznie problematyczne – według Kanta człowiek nigdy nie posiada bezpośredniego poznania rzeczy samej w sobie, także w świecie analogowym. Cyfryzacja nie odbiera więc pierwotnego dostępu do noumenu, którego wcześniej posiadaliśmy. Można jednak wykorzystać Kanta subtelniej: system cyfrowy wprowadza kolejną warstwę warunków możliwego doświadczenia. Człowiek nie poznaje już wyłącznie świata ukształtowanego przez własne formy poznawcze; coraz częściej otrzymuje fenomeny wstępnie wyselekcjonowane, sklasyfikowane i syntetyzowane przez cudze systemy techniczne. To właśnie stanowi rzeczywisty przełom epistemologiczny.

Interfejs nie jest oknem. Jest filtrem posiadającym własną logikę. Mapa wybiera istotne drogi, wyszukiwarka szereguje wyniki, aparat koryguje kolor, algorytm rekomendacyjny porządkuje

treści, a model generatywny tworzy syntezę. W każdym z tych przypadków człowiek otrzymuje rzeczywistość już poddaną transformacji. Problem nie polega na samym zapośredniczeniu – poznanie zawsze było zapośredniczone – lecz na nieprzejrzystości kryteriów zapośredniczenia. Pozwala to powrócić do Korzybskiego.

Zasada „mapa nie jest terytorium" nie oznacza, że z map należy zrezygnować. Dobra mapa bywa poznawczo lepsza niż bezpośrednia obserwacja, gdyż ukazuje strukturę niedostępną z poziomu ulicy. Zagrożenie pojawia się dopiero wtedy, gdy użytkownik traci świadomość selekcyjnego charakteru narzędzia. W erze AI należałoby rozszerzyć tę zasadę: model nie jest światem, odpowiedź nie jest źródłem, prawdopodobieństwo nie jest prawdą, a płynność nie jest dowodem. Kryzys rzeczywistości powinien być zatem traktowany mniej katastroficznie, a bardziej instytucjonalnie. Człowiek nie potrzebuje niemożliwego „powrotu do rzeczy samych" w całkowicie niezapośredniczonej postaci.

Budowa epistemicznej konstytucji świata syntetycznego

Konieczny jest system społeczny umożliwiający kontrolę nad pośrednikami: źródłami, autorami, mediami, modelami i platformami. Ochrona realności nie oznacza odrzucenia reprezentacji, lecz zachowanie **odwracalnej drogi od reprezentacji do jej podstaw**. Na tym właśnie polega normatywne znaczenie proweniencji. Nie chodzi o to, by każda treść syntetyczna była z definicji traktowana jako gorsza; fikcja, rekonstrukcja historyczna, satyra, sztuka czy symulacja naukowa mają przecież ogromną wartość cywilizacyjną.

Wymóg etyczny dotyczy rozpoznawalności statusu reprezentacji. Społeczeństwo musi mieć możliwość odróżnienia dokumentu od inscenizacji, rekonstrukcji od zapisu, a syntezy od świadectwa. Wartością chronioną nie jest tutaj „analogowość", lecz **uczciwość ontologiczna reprezentacji**. Rozróżnienie to ma kluczowe znaczenie również dla wolności sztuki. Całkowity zakaz generowania syntetycznych obrazów byłby nierealistyczny i aksjologicznie ubogi. Chesney i Citron zwracali uwagę, że techniki syntetyczne znajdują zastosowanie w edukacji i sztuce, a ich generalna prohibicja niosłaby za sobą poważne koszty dla wolności ekspresji.

Z perspektywy aksjologicznej nie dążymy zatem do przeciwstawienia sobie „prawdy" i „wyobraźni" – cywilizacja potrzebuje obu. Problem pojawia się w momencie, gdy fikcja podszywa się pod dowód, a odbiorcy odmawia się informacji niezbędnej do rozpoznania tej różnicy. Można zatem sformułować pierwszą dojrzałą zasadę aksjologiczną ery syntetycznej: **im większa zdolność**

reprezentacji do imitowania świadectwa faktualnego, tym większa odpowiedzialność za ujawnienie jej genezy. Nie wynika to z nostalgii za fotografią analogową, lecz z zasady autonomii odbiorcy. Człowiek może świadomie wybrać udział w fikcji, ale nie jest w stanie dokonać takiego wyboru, jeśli ta została przedstawiona jako dokument.

Drugą zasadą jest ochrona integralności tożsamości. Twarz i głos nie są zwykłymi znakami; pełnią funkcję identyfikującą i społeczną, wiążąc osobę z jej działaniami. Technologia zdolna do ich odtworzenia bez udziału podmiotu wkracza zatem w sferę szczególnego dobra osobistego. Najnowsze analizy prawne deepfake'ów słusznie łączą problem detekcji z prawami człowieka, prywatnością, ochroną danych, wolnością ekspresji i niedyskryminacją. Aksjologia nie może tu poprzestać na wartości informacji; musi objąć wartość osoby.

Trzecią zasadą jest proporcjonalność sceptycyzmu. Społeczeństwo wierzące każdemu obrazowi staje się podatne na oszustwo, natomiast takie, które nie wierzy żadnemu, zapada na cynizm i akceptuje bezkarność. Zdolność demokratyczna polega na utrzymaniu przestrzeni pośredniej: wystarczającej otwartości na świadectwo, by współpraca była możliwa, oraz wystarczającej krytyczności, by oszustwo mogło zostać wykryte. Największym zwycięstwem symulakrum nie byłoby stworzenie perfekcyjnej iluzji, lecz przekonanie człowieka, że **nie ma już żadnej różnicy pomiędzy iluzją i faktem**. Dopiero wtedy kryterium prawdy zostaje naprawdę osłabione, gdyż społeczeństwo rezygnuje z samego wysiłku odróżniania. Empiryczne badania nad deepfake'ami sugerują właśnie ten mniej spektakularny, lecz głęboki mechanizm: nie tylko powstawanie fałszywych przekonań, ale wzrost ogólnej niepewności i cynizmu.

Dlatego „pustynia rzeczywistości" nie musi oznaczać świata całkowicie sztucznego. Może opisywać świat, w którym rzeczywistość nadal istnieje, lecz społeczny koszt jej ustalania stał się na tyle wysoki, że coraz mniej podmiotów chce go ponosić. To scenariusz znacznie bardziej realistyczny niż Baudrillardowska metafizyka zniknięcia realnego. Materialny świat nie odchodzi; odchodzić może natomiast **wspólna procedura uzgadniania tego, co wydarzyło się w materialnym świecie**. W takim ujęciu kryzys rzeczywistości staje się kryzysem instytucji prawdy: nauki, dziennikarstwa, sądów, archiwów, edukacji, systemów certyfikacji i odpowiedzialności źródeł. Każda z tych instytucji istnieje po to, by społeczeństwo nie musiało rekonstruować prawdy od podstaw przy każdej informacji. Generatywna AI zwiększa presję na te instytucje, ale może również wzmocnić ich sprawczość, dostarczając nowych narzędzi analizy i wykrywania niezgodności.

Wniosek nie powinien brzmieć: „powróćmy od symulacji do analogowej realności". Bardziej dojrzała teza jest trudniejsza: człowiek zawsze żył jednocześnie w świecie rzeczy i świecie znaków. Cyfryzacja dramatycznie zwiększyła autonomię znaków, a generatywna AI obniżyła koszt produkcji znaków bez referenta. Odpowiedzią cywilizacyjną nie może być wyrzeczenie się reprezentacji, lecz

budowa **epistemicznej konstytucji świata syntetycznego**: norm proweniencji, odpowiedzialności, weryfikowalności, prawa do wiedzy o syntetyczności przekazu oraz instytucji chroniących warunki racjonalnego zaufania. Wartość prawdy spotyka się tutaj z godnością, autonomia z odpowiedzialnością, wolność ekspresji z integralnością osoby, a twórczość z obowiązkiem niepodszywania się pod cudze świadectwo.

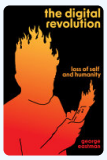
Aksjologia nie jest zatem dekoracyjnym dodatkiem do technologii, lecz narzędziem rozstrzygania konfliktów między wartościami, których nie da się jednocześnie maksymalizować bez ograniczeń. Prawda bez wolności może stać się dogmatem; wolność bez prawdy – manipulowalną arbitralnością; prywatność bez odpowiedzialności może ukrywać nadużycie, a pełna transparentność bez prywatności degradować godność. Dojrzałe społeczeństwo cyfrowe nie znajdzie jednego parametru optymalizacyjnego, który rozwiąże te konflikty. Musi nauczyć się je **instytucjonalnie równoważyć**.

Wpływ symulakrów na ludzką tożsamość i relacje

Ta aksjologiczna trudność stanie się jeszcze bardziej wyraźna, gdy problem symulakrum przejdzie w problem tożsamości, relacji i kultury. Trzeba bowiem zapytać nie tylko o to, czy cyfrowy obraz wiernie przedstawia świat, lecz także, jak długotrwałe życie pośród reprezentacji zmienia sposób, w jaki człowiek przedstawia samego siebie, buduje więzi, rozumie autentyczność i negocjuje granicę pomiędzy prywatnym "ja" a publicznie mierzalnym profilem.

> Być może najgłębszym zwycięstwem symulakrum nie będzie stworzenie perfekcyjnej iluzji, lecz przekonanie nas, że różnica między faktem a fikcją przestała mieć znaczenie. W świecie, gdzie koszt produkcji pozoru dąży do zera, prawdziwa wolność przestaje być brakiem przymusu, a staje się rzadką zdolnością odróżnienia mapy od terytorium. Czy w obliczu tej asymetrii będziemy w stanie zbudować nową konstytucję prawdy, czy też zaakceptujemy rolę pasywnych odbiorców w rzeczywistości, która nie potrzebuje już oryginału?

Inspiracje

[1] 

"The Digital Revolution Loss of Humanity" George Eastman

ISBN: 9781945335655

Dr George H. Eastman jest amerykańskim psychologiem klinicznym, filozofem i pedagogiem. Uzyskał tytuł doktora edukacji na Uniwersytecie Harvarda w 1963 roku, specjalizując się w filozofii, językoznawstwie i relacjach społecznych, a następnie, w 1974 roku, uzyskał doktorat z psychologii klinicznej na Uniwersytecie Nowojorskim, koncentrując się na teoriach gier i podejmowania decyzji. Eastman pełnił funkcję profesora zwyczajnego w systemie Uniwersytetu Stanowego Nowego Jorku i przez dwadzieścia cztery lata wykładał muzykoterapię w Berklee College of Music, a także przez kilkadziesiąt lat prowadził prywatną praktykę psychoterapeutyczną w Cambridge w stanie Massachusetts. Założył również organizację non-profit Institute of Media Research, aby badać fizjologiczny i psychologiczny wpływ nowoczesnych mediów cyfrowych na rozwój poznawczy i wolność człowieka. Jego publikacje zgłębiają autonomię człowieka, psychoanalizę oraz wpływ systemów technologicznych na tożsamość jednostki i relacje interpersonalne.

Dr. George H. Eastman is an American clinical psychologist, philosopher, and educator. He earned a Doctor of Education from Harvard University in 1963, specializing in philosophy, linguistics, and social relations, and later completed a Ph.D. in Clinical Psychology at New York University in 1974, focusing on game and decision-making theories. Eastman served as a tenured professor in the State University of New York system and taught music therapists at Berklee College of Music for twenty-four years, in addition to maintaining a private psychotherapy practice in Cambridge, Massachusetts, for several decades. He also established the nonprofit Institute of Media Research to examine the physiological and psychological effects of modern digital media on cognitive development and human freedom. His published writings explore human autonomy, psychoanalysis, and the impact of technological systems on individual identity and interpersonal relationships.

Institute of Media Research

Literatura uzupełniająca

Książki

Książki autora i dzieła przywołane w tekście:



[1] "Chronicles of an African Trip"

George Eastman

2018 | Papamoa Press | ISBN: 9781789126112



[2] "The Reigns of King Solomon and Queen Victoria. A Sermon [on 1 Kings I. 38-40] Preached ... June 20, 1869 ... the Anniversary of Her Majesty's Accession"

Rev. George EASTMAN

1869



[3] "Getting Started in Clinical Radiology"

George W. Eastman, Christoph Wald

2011 | Thieme | ISBN: 9781604061154



[4] "Chronicles of a Second African Trip"

George Eastman

1987 | University of Rochester Libraries

Książki autorów przywoływanych i literatura pokrewna:



[5] "The Mystery of Consciousness"

John Searle

1997 | ISBN: 9781862070745



[6] "Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics"

Alfred Korzybski

Institute of GS | ISBN: 9780937298015



[7] "Mind: A Brief Introduction"

John Searle

Oxford University Press | ISBN: 9780199729005



[8] "Seeing Things as They Are: A Theory of Perception"

John Searle

Oxford University Press, USA | ISBN: 9780199385157



[9] "Open Access Archivangelism"

Stevan Harnad

American Library Association | ISBN: 9780838911068



[10] "Reason, Truth and History"

Hilary Putnam

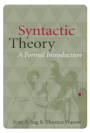
1991 | ISBN: 9780521230353



[11] "Ethics Without Ontology"

Hilary Putnam

2004 | Harvard University Press | ISBN: 9780674013100



[12] "Syntactic Theory: A Formal Introduction"

Emily Bender

Stanford Univ Center for the Study | ISBN: 9781575861609



[13] "Black in AI"

Timnit Gebru

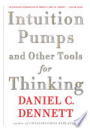
Berrett-Koehler Publishers | ISBN: 9798890571908



[14] "Consciousness Explained"

Daniel C. Dennett

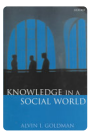
2017 | Little, Brown | ISBN: 9780316439480



[15] "Intuition Pumps And Other Tools for Thinking"

Daniel C. Dennett

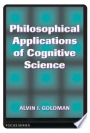
2013 | W. W. Norton & Company | ISBN: 9780393240689



[16] "Knowledge in a Social World"

Alvin I. Goldman, Board of Governors Professor of Philosophy and Cognitive Science Alvin I Goldman

1999 | Oxford University Press on Demand | ISBN: 9780198237778



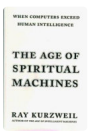
[17] "Philosophical Applications Of Cognitive Science"

Alvin I. Goldman

2018 | Routledge | ISBN: 9780429972980

[18] "Claude Lévi-Strauss"

Dan Sperber



[19] "The Age of Spiritual Machines"

Ray Kurzweil

1999 | Viking Adult



[20] "The Age of Intelligent Machines"

Ray Kurzweil

1990 | MIT Press (MA)



[21] "Human Enhancement"

Nick Bostrom

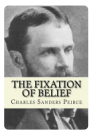
OUP Oxford | ISBN: 9780191559600



[22] "A Neglected Argument for the Reality of God"

Charles Sanders Peirce
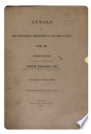
Edizioni Aurora Boreale | ISBN: 9791255046585



[23] "The Fixation of Belief"

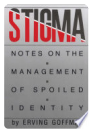
Charles Sanders Peirce

Createspace Independent Publishing Platform | ISBN: 9781973922995



[24] "Photometric Researches"

Charles Sanders Peirce



[25] "Stigma: Notes on the Management of Spoiled Identity"

Erving Goffman

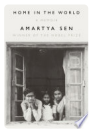
Simon and Schuster | ISBN: 9781439188330



[26] "Identity and Violence"

Amartya Sen

W. W. Norton & Company | ISBN: 9780393243192



[27] "Home in the World"

Amartya Sen

Liveright Publishing | ISBN: 9781324091622



[28] "Epistemic Injustice: Power and the Ethics of Knowing"

Miranda Fricker

Oxford University Press on Demand | ISBN: 9780199570522



[29] "Epistemic Injustice"

Miranda Fricker

2007 | Oxford University Press | ISBN: 9780198237907



[30] **"On Truth"**

Harry Frankfurt

2010 | ISBN: 9781407064598



[31] **"On Bullshit"**

Harry Frankfurt

2025 | Princeton University Press | ISBN: 9780691276793

Artykuły naukowe

Prace autorów przywoływanych w tekście:

[1] **"Social epistemics and social psychology"**

Alvin Goldman

1991 | Social Epistemology | DOI: 10.1080/02691729108578607

Źródło

[2] **"Speech Acts: An Essay in the Philosophy of Language."**

William P. Alston, John Rogers Searle

1970 | The Philosophical Quarterly | DOI: 10.2307/2218090

Źródło

[3] **"Intentionality: Name index"**

John Rogers Searle

1983

Źródło

[4] **"Intentionality: Subject index"**

John Rogers Searle

1983

Źródło

[5] **"Making the Social World"**

John Rogers Searle

2010 | Oxford University Press eBooks | DOI: 10.1093/acprof:osobl/9780195396171.001.0001

Źródło

[6] **"Can Digital Computers Think? (1951)"**

Alan Turing

2004 | The Essential Turing | DOI: 10.1093/oso/9780198250791.003.0019

Źródło

[7] **"Can Automatic Calculating Machines Be Said To Think? (1952)"**

Alan Turing, Richard Braithwaite

2004 | The Essential Turing | DOI: 10.1093/oso/9780198250791.003.0020

Źródło

[8] **"Machine Learning, a key component in business model transformation"**

A. Turing

2018

Źródło

[9] **"Reason, Truth and History"**

Hilary Putnam

1981 | Cambridge University Press eBooks | DOI: 10.1017/cbo9780511625398

Źródło

[10] **"Representation and Reality"**

Hilary Putnam

1988 | The MIT Press eBooks | DOI: 10.7551/mitpress/5891.001.0001

Źródło

[11] **"A Conversation between Jacques Bouveresse and Hilary Putnam"**

J. Bouveresse, H. Putnam

2020 | DOI: 10.1093/monist/onaa019

Źródło

[12] **"Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data"**

Emily M. Bender, Alexander Koller

2020 | Proceedings of the conference - Association for Computational Linguistics. Meeting | DOI: 10.18653/v1/2020.acl-main.463

Źródło

[13] "Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science"

Emily M. Bender, Batya Friedman

2018 | Transactions of the Association for Computational Linguistics | DOI: 10.1162/tacl_a_00041

Źródło

[14] "Data and its (dis)contents: A survey of dataset development and use in machine learning research"

Amandalynne Paullada, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, Alex Hanna

2021 | Patterns | DOI: 10.1016/j.patter.2021.100336

Źródło

[15] "AI and the Everything in the Whole Wide World Benchmark"

Inioluwa Deborah Raji, Emily M. Bender, Amandalynne Paullada, Emily Denton, Alex Hanna

2021 | arXiv (Cornell University) | DOI: 10.48550/arxiv.2111.15366

Źródło

[16] "The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence"

Timnit Gebru, Émile P. Torres

2024 | First Monday | DOI: 10.5210/fm.v29i4.13636

Źródło

[17] "What is Your Favorite Gender, MLM? Gender Bias Evaluation in Multilingual Masked Language Models"

Jinho D. Choi, T. Gebru, Angelina McMillan-642, Su Lin Blodgett, S. Barocas

2024 | Inf. | DOI: 10.48550/arxiv.2404.06621

Źródło

[18] "Community Driven Approaches to Research in Technology & Society CCC Workshop Report"

S. Venkatasubramanian, T. Gebru, U. Topcu, Haley Griffin, Leah Namisa Rosenbloom

2024 | arXiv.org | DOI: 10.48550/arxiv.2406.07556

Źródło

[19] **"Documenting Geographically and Contextually Diverse Data Sources: The BigScience Catalogue of Language Data and Resources"**

Angelina McMillan-Major, Zaid Alyafeai, Stella Biderman, Kimbo Chen, Francesco De Toni

2022 | arXiv (Cornell University) | DOI: 10.48550/arxiv.2201.10066

Źródło

[20] **"An Interactive Exploratory Tool for the Task of Hate Speech Detection"**

Angelina McMillan-Major, Amandalynne Paullada, Yacine Jernite

2022 | Workshop on Bridging Human-Computer Interaction and Natural Language Processing (HCI+NLP) | DOI: 10.18653/v1/2022.hcinlp-1.2

Źródło

[21] **"3 A Critical Taxonomic Analysis of Serious Games and Language Learning"**

Angelina McMillan-Major, Russell Hugo

2026 | University of Toronto Press eBooks | DOI: 10.3138/9781049805177-005

Źródło

[22] **"Documenting Geographically and Contextually Diverse Language Data Sources"**

Angelina McMillan-Major, Francesco De Toni, Zaid Alyafeai, Stella Biderman, Kimbo Chen

2024 | Northern European Journal of Language Technology | DOI: 10.3384/nejlt.2000-1533.2024.5217

Źródło

[23] **"Midge: Generating Image Descriptions From Computer Vision Detections"**

Margaret Mitchell, Jesse Dodge, Amit Goyal, Kota Yamaguchi, Karl Stratos

2012

Źródło

[24] **"Book Reviews: Sociology: The State of the Art, A Treatise on Social Theory, Volume 1: The Methodology of Social Theory, Contemporary Political Philosophy: Radical Studies, Margaret Mead and Samoa: The Making and Unmaking of an Anthropological Myth, Capital and Politics, Supremacy and Subordination of Labour; The Hierarchy of Work in the Early Labour Movement, The Development of the Labour Process in Capitalist Societies; A Comparative Study of the Transformation of Work Organisation in Britain, Japan and the USA, Blue-Collar Workers and Politics: A French Paradox, Contemporary French Political Parties, The Family and Industrial Society, White-Collar Work, Industrial Relations in Britain, Women in Control: Dilemmas of a Workers Co-Operative, An Economic History of Women in America, Racial and Ethnic Competition, Labour and Racism, Racism and Migrant Labour, Contemporary Education Policy, Solidarity: The Analysis of a Social Movement: Poland 1980–1981, The Rites of Rulers, Political Anatomy of the Body; Medical Knowledge in Britain in the Twentieth Century, Class, Sports, and Social Development, Cambridge College of Arts and Technology Orality and Literacy: The Technologizing of the Word, From Author to Reader"**

Duncan Mitchell, David Jary, Kenneth Thompson, Keith Tribe, Michael Winter

1984 | The Sociological Review | DOI: 10.1111/j.1467-954x.1984.tb00821.x

Źródło

[25] **"Dissociating language and thought in large language models"**

Kyle Mahowald, Anna Aleksandrovna Ivanova, Idan Blank, Nancy Kanwisher, Joshua B. Tenenbaum

2024 | Trends in Cognitive Sciences | DOI: 10.1016/j.tics.2024.01.011

Źródło

[26] **"How Efficiency Shapes Human Language"**

Edward A. F. Gibson, Richard Futrell, Steven T. Piantadosi, Isabelle Dautriche, Kyle Mahowald

2019 | Trends in Cognitive Sciences | DOI: 10.1016/j.tics.2019.02.003

Źródło

[27] **"Large-scale evidence of dependency length minimization in 37 languages"**

Richard Futrell, Kyle Mahowald, Edward A. F. Gibson

2015 | Proceedings of the National Academy of Sciences | DOI: 10.1073/pnas.1502134112

Źródło

[28] **"A meta-analysis of syntactic priming in language production"**

Kyle Mahowald, Ariel N. James, Richard Futrell, Edward A. F. Gibson

2016 | Journal of Memory and Language | DOI: 10.1016/j.jml.2016.03.009

Źródło

[29] "Google's powerful AI spotlights a human cognitive glitch: Mistaking fluent speech for fluent thought"

Kyle Mahowald, Anna Ivanova

2022 | DOI: 10.64628/aai.wvs9mp5dy

Źródło

[30] "PRINCIPLES OF PUBLIC ADMINISTRATION OF SCIENTIFIC RESEARCHES USING ARTIFICIAL INTELLIGENCE"

Anna Ivanova

2025 | Drukerovskij Vestnik | DOI: 10.17213/2312-6469-2025-6-43-51

Źródło

[31] "The neural architecture of language: Integrative modeling converges on predictive processing"

Martin Schrimpf, Idan Blank, Greta Tuckute, Carina Kauf, Eghbal A. Hosseini

2021 | Proceedings of the National Academy of Sciences | DOI: 10.1073/pnas.2105646118

Źródło

[32] "Semantic projection recovers rich human knowledge of multiple object features from word embeddings"

Gabriel Grand, Idan Blank, Francisco Pereira, Evelina G. Fedorenko

2022 | Nature Human Behaviour | DOI: 10.1038/s41562-022-01316-8

Źródło

[33] "fMRI reveals language-specific predictive coding during naturalistic sentence comprehension"

Cory Shain, Idan Blank, Marten van Schijndel, William Schuler, Evelina G. Fedorenko

2019 | Neuropsychologia | DOI: 10.1016/j.neuropsychologia.2019.107307

Źródło

[34] "The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception"

Nancy Kanwisher, Josh H. McDermott, Marvin M. Chun

1997 | Journal of Neuroscience | DOI: 10.1523/jneurosci.17-11-04302.1997

Źródło

[35] "A cortical representation of the local visual environment"

Russell A. Epstein, Nancy Kanwisher

1998 | Nature | DOI: 10.1038/33402

Źródło

[36] "A Cortical Area Selective for Visual Processing of the Human Body"

Paul E. Downing, Yuhong V. Jiang, Miles Shuman, Nancy Kanwisher

2001 | Science | DOI: 10.1126/science.1063414

Źródło

[37] "Functional specificity in the human brain: A window into the functional architecture of the mind"

Nancy Kanwisher

2010 | Proceedings of the National Academy of Sciences | DOI: 10.1073/pnas.1005062107

Źródło

[38] "Human-level concept learning through probabilistic program induction"

Brenden M. Lake, Ruslan Salakhutdinov, Joshua B. Tenenbaum

2015 | Science | DOI: 10.1126/science.aab3050

Źródło

[39] "Building machines that learn and think like people"

Brenden M. Lake, Tomer David Ullman, Joshua B. Tenenbaum, Samuel Joseph Gershman

2016 | Behavioral and Brain Sciences | DOI: 10.1017/s0140525x16001837

Źródło

[40] "Machine behaviour"

Iyad Rahwan, Manuel Cebrián, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Joshua B. Tenenbaum

2019 | Nature | DOI: 10.1038/s41586-019-1138-y

Źródło

[41] "Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling"

Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T. Freeman, Joshua B. Tenenbaum

2016 | arXiv (Cornell University) | DOI: 10.48550/arxiv.1610.07584

Źródło

[42] "New Method for fMRI Investigations of Language: Defining ROIs Functionally in Individual Subjects"

Evelina G. Fedorenko, Po-Jang Hsieh, Alfonso Nieto-Castañón, Susan Whitfield, Nancy Kanwisher

2010 | Journal of Neurophysiology | DOI: 10.1152/jn.00032.2010

Źródło

[43] "Reworking the language network"

Evelina G. Fedorenko, Sharon L. Thompson-Schill

2014 | Trends in Cognitive Sciences | DOI: 10.1016/j.tics.2013.12.006

Źródło

[44] "Functional specificity for high-level linguistic processing in the human brain"

Evelina G. Fedorenko, Michael K. Behr, Nancy Kanwisher

2011 | Proceedings of the National Academy of Sciences | DOI: 10.1073/pnas.1112937108

Źródło

[45] "Get Real"

Daniel C. Dennett

1994 | Philosophical Topics | DOI: 10.5840/philtopics1994221/212

Źródło

[46] "From Typo to Thinko: When Evolution Graduated to Semantic Norms"

Daniel C. Dennett

2005 | The MIT Press eBooks | DOI: 10.7551/mitpress/2870.003.0010

Źródło

[47] "Human Rationality: Epistemological and Psychological Perspectives"

Alvin Goldman

2008 | Philosophie: Grundlagen und Anwendungen /Philosophy: Foundations and Applications | DOI: 10.30965/9783969750056_017

Źródło

[48] "Why do humans reason? Arguments for an argumentative theory"

Hugo Mercier, Dan Sperber

2011 | Behavioral and Brain Sciences | DOI: 10.1017/s0140525x10000968

Źródło

[49] "Epistemic Vigilance"

Dan Sperber, Fabrice Clément, Christophe Heintz, Olivier Mascaro, Hugo Mercier

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

Źródło

[50] "Anthropology and Psychology: Towards an Epidemiology of Representations"

Dan Sperber

1985 | Man | DOI: 10.2307/2802222

Źródło

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

Olivier Mascaro, Dan Sperber

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

Źródło

[52] "The Singularity Is near: When Humans Transcend Biology"

Walter Russell Mead, Ray Kurzweil

2006 | Foreign Affairs | DOI: 10.2307/20031996

Źródło

[53] "The age of spiritual machines: when computers exceed human intelligence"

Ray Kurzweil

1998

Źródło

[54] "The Age of Spiritual Machines"

Ray Kurzweil

1999

Źródło

[55] "Efficient Natural Language Response Suggestion for Smart Reply"

Matthew Henderson, Rami Al-Rfou, Brian P. Strope, Yun-Hsuan Sung, Lukacs, Laszlo, Ray Kurzweil

2017 | arXiv (Cornell University) | DOI: 10.48550/arxiv.1705.00652

Źródło

[56] "Magritte, ideas and images"

René Magritte, Harry Torczyner

1977

Źródło

[57] "Magritte: The silence of the world"

David Sylvester, René Magritte

1969

Źródło

[58] "Magritte and Contemporary Art: The Treachery of Images"

René Magritte, Stephanie Barron, Michel Draguet, Sara Cochran

2006 | Dépôt institutionnel de l'Université libre de Bruxelles (Université Libre de Bruxelles)

Źródło

[59] "The Human Condition"

René Magritte

2007 | CSUN ScholarWorks (California State University, Northridge)

Źródło

[60] "Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics."

Howard S. Becker, Alfred Korzybski

1942 | American Sociological Review | DOI: 10.2307/2085182

Źródło

[61] "Manhood of Humanity"

Alfred Korzybski

2021

Źródło

[62] "Science and sanity : an introduction to non-aristotelian systems and general semantics / Alfred Korzybski"

Alfred Korzybski

1958

Źródło

[63] **"The role of language in the perceptual process"**

Alfred Korzybski

1951 | MPG.PuRe (Max Planck Society)

Źródło

[64] **"Language. An Introduction to the Study of Speech"**

W. E. Collinson, Edward Sapir, Ferdinand de Saussure

1924 | The Modern Language Review | DOI: 10.2307/3713880

Źródło

[65] **"Les sources manuscrites du Cours de linguistique générale de F. de Saussure"**

Robert Godel, Ferdinand de Saussure

1957

Źródło

[66] **"Refutación de todos los juicios, tanto elogiosos como hostiles, que hasta ahora se han emitido sobre la película la sociedad del espectáculo"**

Guy Debord

1999 | RiuNet (Universitat Politècnica de València)

Źródło

[67] **"Stigma: Notes on the Management of Spoiled Identity."**

Melvin S Seeman, Erving Goffman

1964 | American Sociological Review | DOI: 10.2307/2091442

Źródło

[68] **"STIGMA: NOTES ON THE MANAGEMENT OF SPOILED IDENTITY. By Erving Goffman. Englewood Cliffs, New Jersey: Prentice-Hall, 1963. 147 pp. Cloth, $4.50; paper, $1.95"**

Melvin L. DeFleur, Erving Goffman

1964 | Social Forces | DOI: 10.1093/sf/43.1.127

Źródło

[69] **"Truth and Truthfulness: An Essay in Genealogy"**

Bernard Williams

2002 | Project Muse (Johns Hopkins University) | DOI: 10.1515/9781400825141

Źródło

[70] "Making Sense of Humanity"

Bernard Williams

1995 | Cambridge University Press eBooks | DOI: 10.1017/cbo9780511621246

Źródło

[71] "Aristotle on the Good: A Formal Sketch"

Bernard Williams

1962 | The Philosophical Quarterly | DOI: 10.2307/2217421

Źródło

[72] "Truth and Truthfulness"

Maurice Meilleur, Bernard Williams

2003 | The Antioch Review | DOI: 10.2307/4614536

Źródło

[73] "Conditionally immortal neuroepithelial stem cell grafts reverse age-associated memory impairments in rats"

H Hodges, T Veizovic, Nicholas J. Bray, Sarah Jane French, Tahira Rashid, Andrew J Chadwick

2000 | Neuroscience | DOI: 10.1016/s0306-4522(00)00408-5

Źródło

[74] "Functional Reconstruction of the Hippocampus: Fetal Versus Conditionally Immortal Neuroepithelial Stem Cell Grafts"

Helen Ruth Hodges, P. Sowinski, David J. Virley, A. Nelson, Tim R. Kershaw, Andrew J Chadwick

2000 | Novartis Foundation symposium | DOI: 10.1002/0470870834.ch4

Źródło

[75] "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News"

Cristian Vaccari, Andrew Chadwick

2020 | Social Media + Society | DOI: 10.1177/2056305120903408

Źródło

[76] "China has more fans at this World Cup than England"

Simon Chadwick

2018 | DOI: 10.64628/ab.gm5npuatp

Źródło

[77] "Apuntes sobre los orígenes de la política en Hannah Arendt"

H. Arendt, Sergio Quintero Martín

2023 | Theoría. Revista del Colegio de Filosofía | DOI: 10.22201/ffyl.29544270e.2023.44.1803

Źródło

[78] "Hannah Arendt: el problema de la responsabilidad ante los crímenes de lesa humanidad en los regímenes totalitarios"

H. Arendt, Yuliana Leal, H. Arendt

2018

Źródło

[79] "Alternate Possibilities and Moral Responsibility"

Harry G. Frankfurt

1969 | The Journal of Philosophy | DOI: 10.2307/2023833

Źródło

[80] "Descartes on the Creation of the Eternal Truths"

Harry G. Frankfurt

1977 | The Philosophical Review | DOI: 10.2307/2184161

Źródło

[81] "On Truth, Lies, and Bullshit"

Harry Frankfurt

2009 | The Philosophy of Deception | DOI: 10.1093/acprof:oso/9780195327939.003.0003

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: c8bafa7f-770c-4901-8eac-12eb27318479