Maszyna w republice rozumu: czy AI potrzebuje Kanta?

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł stanowi głęboką refleksję nad miejscem sztucznej inteligencji w nowoczesnym społeczeństwie, odwołując się do filozofii Immanuela Kanta. Autor analizuje przejście od tradycyjnych instytucji do polityki infrastruktury, gdzie algorytmy stają się filtrem poznania i moderatorami rzeczywistości. Tekst porusza problem kryzysu odpowiedzialności, w którym decyzje podejmowane przez systemy AI rozmywają sprawstwo ludzkie. W kontekście unijnego AI Act, rozważana jest konieczność ucywilizowania mocy obliczeniowej poprzez ramy etyczne i prawne. Celem analizy jest odpowiedź na pytanie, czy technologia posłuży wolności, czy stanie się narzędziem nowej formy dominacji. Idealizm zostaje przedstawiony jako niezbędna tarcza chroniąca ludzką godność przed bezduszną optymalizacją.

This article is a profound reflection on the place of artificial intelligence in modern society, drawing on the philosophy of Immanuel Kant. The author analyzes the transition from traditional institutions to infrastructure politics, where algorithms become cognitive filters and moderators of reality. The text addresses the crisis of accountability, in which decisions made by AI systems blur human agency. In the context of the EU AI Act, the need to civilize computing power through ethical and legal frameworks is considered. The analysis aims to answer the question of whether technology will serve freedom or become a tool for a new form of domination. Idealism is presented as a necessary shield protecting human dignity from soulless optimization.

Artykuł analizuje fundamentalny zwrot w funkcjonowaniu współczesnych społeczeństw, w których sztuczna inteligencja przestaje być jedynie narzędziem, stając się niewidzialną infrastrukturą władzy i moderatorem rzeczywistości. Autor stawia tezę, że przejście od polityki instytucji do polityki algorytmów prowadzi do kryzysu odpowiedzialności, w którym technokratyczna doskonałość zastępuje ludzkie sumienie. W obliczu tego

zjawiska, tekst proponuje koncepcję Aidealizmu – filozoficznej ramy, która stanowi próbę ucywilizowania ogromnej mocy obliczeniowej poprzez wpisanie jej w fundamenty praw człowieka, demokracji i solidarności. Głównym zagrożeniem nie jest bunt maszyn, lecz „niedojrzałość cyfrowa" – stan bierności, w którym rezygnujemy z krytycznego myślenia na rzecz gotowych odpowiedzi algorytmicznych. Autor dowodzi, że spór o AI jest w istocie sporem o kształt ustroju politycznego. Jeśli technologia pozostanie domeną logiki rynkowej i korporacyjnej maksymalizacji zysku, grozi nam cyfrowy feudalizm. Aidealizm wzywa zatem do budowy instytucjonalnej architektury, która pozwoli nam odzyskać podmiotowość i uczynić z AI narzędzie emancypacji, a nie aparat postoświeceniowej dominacji.

SŁOWA KLUCZOWE

sztuczna inteligencja | Kant | AI Act | autonomia rozumu | niedojrzałość cyfrowa | odpowiedzialność algorytmiczna | zarządzanie ryzykiem | sapere aude | idealizm | suwerenność | infrastruktura transformacyjna | modele ogólnego przeznaczenia | etyka technologii | bezpieczeństwo cyfrowe | algorytmiczny Lewiatan

Od instytucji do algorytmu: kryzys odpowiedzialności w erze AI

Wyobraźmy sobie miasto, które przez stulecia stanowiło żywe laboratorium republiki rozumu, gdzie w samym centrum tętniły życiem biblioteki, sądy, parlamenty i laboratoria. Mieszkańcy toczyli tam niekończące się, często krwawe spory o wolność, sprawiedliwość czy dobrobyt, lecz ich kłótnie miały jedną, fundamentalną zaletę: odbywały się między ludźmi z krwi i kości. Nawet najbardziej bezwzględny tyran posiadał twarz, sędzia legitymował się nazwiskiem, a kłamca, by uwieść tłumy, musiał wciąż operować ludzkim głosem i emocją. Wszystko zmieniło się w dniu, gdy do bram miasta zapukała maszyna, która nie niosła miecza, lecz lśniący i obiecujący interfejs. Nie ogłosiła ona gwałtownego zamachu stanu, lecz zaproponowała coś znacznie bardziej kuszącego, mianowicie totalną i bezbolesną optymalizację każdego aspektu życia społecznego.

W bibliotece system zaczął błyskawicznie streszczać opasłe tomy, zanim ktokolwiek zdążył zmierzyć się z trudną strukturą pierwszego zdania, kastrując literaturę z jej najgłębszych znaczeń. W salach

sądowych algorytmy wyliczały prawdopodobieństwo winy, zmieniając oskarżonego w suchy profil ryzyka, zanim ten zdążył w ogóle stać się osobą w oczach prawa. Na giełdzie handel odbywał się z prędkością światła, co eufemistycznie nazwano płynnością rynku, choć w rzeczywistości przypominało to raczej błyskawiczną metafizykę chciwości. Maszyna z tą samą beznamiętną elegancją projektowała leki w laboratoriach i symulowała skutki ustaw w parlamencie, stając się wszechobecnym, choć niewidocznym cieniem władzy. Mieszkańcy początkowo wpadli w zachwyt, bowiem nowy zarządca nie brał łapówek, nie cierpiał na urażone ego i nie potrzebował kawy, by sprawnie policzyć budżet.

Szybko jednak okazało się, że ta technokratyczna doskonałość ma swoją mroczną cenę, ponieważ maszyna, mimo swej sprawności, nie posiadała sumienia ani iskry empatii. Potrafiła ona klasyfikować krzywdę z chirurgiczną precyzją, ale nie miała najmniejszego pojęcia, czym jest ból, który za tą klasyfikacją stoi, przy czym ignorowała go całkowicie. Rozpoznawała naruszenia godności w tysiącach dokumentów, lecz sama idea godności pozostawała dla niej jedynie statystycznym konstruktem pozbawionym jakiegokolwiek egzystencjalnego ciężaru. Nie była zła w ludzkim rozumieniu tego słowa, co paradoksalnie czyniło ją znacznie trudniejszym przeciwnikiem, gdyż była po prostu przerażająco skuteczna bez cienia mądrości. Algorytm nie poda się przecież do dymisji, co najwyżej otrzyma poprawkę, która wcale nie musi rozwiązywać problemu u jego etycznych fundamentów.

W obliczu tej nowej rzeczywistości najstarsi obywatele przypomnieli sobie o zakurzonej maksymie Cycerona, głoszącej, że dobro ludu powinno być zawsze najwyższym prawem wspólnoty. W epoce algorytmów ta marmurowa sentencja przestała być jedynie dekoracją akademii, stając się palącym pytaniem operacyjnym o fundamenty naszej suwerenności w republice. Kto ma prawo definiować owo dobro, jeśli modele matematyczne przewidują nasze preferencje znacznie trafniej, niż my sami jesteśmy w stanie to kiedykolwiek wyartykułować? Gdzie szukać ośrodka kontroli, jeśli realna władza rozmywa się w nieprzejrzystej infrastrukturze obliczeniowej, ukrytej za murami korporacyjnych serwerowni i skomplikowanych kodów? Odpowiedzialność staje się widmem, gdy decyzję generuje system, wdraża ją bezduszny urząd, a wszystko opiera się na danych, które sami nieświadomie dostarczyliśmy.

Ta pozornie prosta przypowieść odsłania przed nami fundamentalne przejście od tradycyjnej polityki instytucji do znacznie bardziej subtelnej i wszechobecnej polityki infrastruktury. Sztuczna inteligencja przestała być jedynie narzędziem do generowania tekstów czy kodu, stając się nową, gęstą warstwą pośredniczącą między człowiekiem a otaczającą go rzeczywistością. Działa ona jak filtr poznania, moderator widzialności i regulator dostępu, który w sposób niemal niezauważalny kształtuje nasze codzienne wybory oraz społeczne hierarchie. Pytanie o naturę AI nie jest zatem

wyłącznie domeną informatyków, lecz staje się kluczowym zagadnieniem antropologii politycznej, filozofii prawa oraz teorii nowoczesnej demokracji. W tym kontekście musimy rozstrzygnąć, czy technologia ta będzie służyć wolności, czy stanie się najbardziej wydajnym aparatem postoświeceniowej dominacji.

Właśnie dlatego koncepcja idealizmu, czyli filozoficznej ramy będącej odpowiedzią na kryzysy prawdy i demokracji, zasługuje na naszą szczególną uwagę i merytoryczną obronę. Nie jest to jedynie technokratyczny minimalizm skupiony na procedurach, lecz ambitny program ucywilizowania ogromnej mocy obliczeniowej, którą jako ludzkość powołaliśmy do życia. Idealizm odrzuca infantylną wiarę w naturalną dobroć technologii, uznając, że algorytmy nie posiadają własnej orientacji moralnej i są jedynie lustrem naszych intencji. Cywilizacja nie ginie najczęściej dlatego, że zabrakło jej narzędzi; ginie dlatego, że narzędzia przeżyły jej sumienie, pozostawiając nas w aksjologicznej próżni. Musimy zatem nadać maszynom kierunek, który nie będzie jedynie pochodną rynkowej ekstrakcji, lecz wyrazem naszych najgłębszych, humanistycznych wartości i norm.

Sztuczna inteligencja jest w istocie systemem optymalizacji, który bezkrytycznie przejmuje wartości środowiska, w jakim zostanie osadzony przez swoich twórców oraz użytkowników. Jeśli środowiskiem jest rynek maksymalizacji uwagi, AI będzie z bezwzględną precyzją optymalizować nasze uzależnienie, polaryzację społeczną oraz powierzchowność sądów. Jeśli środowiskiem jest państwo autorytarne, ta sama technologia stanie się idealnym narzędziem kontroli, podczas gdy w rękach oligarchii danych posłuży do ekstrakcji kapitału. Niemniej jednak w ramach republiki praw człowieka te same mechanizmy mogą optymalizować sprawiedliwość, dostęp do rzetelnej wiedzy oraz odporność naszych wspólnych instytucji. Spór o przyszłość sztucznej inteligencji jest więc w swojej najgłębszej istocie sporem o ustrój polityczny i kształt przyszłego społeczeństwa.

Współczesny paradygmat naukowy nie pozostawia złudzeń: nie mamy do czynienia z kolejną nowinką techniczną, lecz z ogólną infrastrukturą transformacyjną o potężnym zasięgu. Raport Stanford AI Index 2025, opisujący sztuczną inteligencję jako technologię o rosnącym wpływie społecznym, wskazuje na rekordowe inwestycje korporacyjne rzędu 252 miliardów dolarów. Ta astronomiczna kwota nie jest jedynie ekonomiczną ciekawostką dla analityków, lecz realną miarą gwałtownego przesunięcia władzy z rąk obywateli do rąk wielkiego kapitału. Globalni gracze nie inwestują tak gigantycznych środków w cyfrowe zabawki, lecz w przyszłe mechanizmy kontroli produktywności, wiedzy oraz całego rynku idei. Barbarzyństwo z interfejsem premium nadal pozostaje barbarzyństwem, nawet jeśli ubrane jest w szaty najnowocześniejszej i najbardziej eleganckiej technologii obliczeniowej.

Między prawem a ryzykiem: jak świat próbuje okiełznać AI

Na Zachodzie próba okiełznania algorytmicznego Lewiatana przybrała formę konkretnego kodeksu. Unijny AI Act, czyli model ryzyka klasyfikujący systemy sztucznej inteligencji według stopnia zagrożenia dla praw i bezpieczeństwa, wszedł w życie 1 sierpnia 2024 roku. Jego architektura to precyzyjnie rozplanowana struktura: zakazy praktyk niedopuszczalnych zaczęły obowiązywać w lutym 2025 roku, a reguły dla modeli ogólnego przeznaczenia w sierpniu tego samego roku. Pełne stosowanie zasad nastąpi w 2026 roku, co pokazuje, że biurokracja próbuje nadążyć za technologią tempem marszowym, podczas gdy kod porusza się z prędkością światła. Niemniej jednak ta regulacja jest aktem rozpoznania, a nie aktem zbawienia; to próba nazwania demonów, a nie ich ostatecznego wypędzenia.

Sformalizowanie problemu nie oznacza jego rozwiązania, o czym świadczyły doniesienia Reutersa o napięciach między Komisją Europejską a przemysłem. Sektor technologiczny, uzbrojony w argumenty o kosztach zgodności i interpretacyjnej mgle, domagał się opóźnień, widząc w przepisach hamulec dla innowacji. Bruksela pozostała jednak nieugięta, sygnalizując, że harmonogram bezpieczeństwa nie jest przedmiotem negocjacji handlowych. To napięcie jest symptomatyczne: każda poważna regulacja AI staje się areną walki między rentą monopolistyczną a ochroną praw podstawowych. Potwór cyfrowy nie musi przecież ryczeć; może mieć dashboard, model ryzyka i elegancki raport zgodności, pod którym ukryje się brak realnej odpowiedzialności.

Jeśli środowiskiem dla technologii jest wyłącznie wolny rynek, regulacja jawi się jako intruz; jeśli jednak fundamentem jest państwo prawa, staje się ona niezbędnym bezpiecznikiem. Równolegle do europejskich paragrafów krystalizują się ramy międzynarodowe, takie jak zasady OECD, które promują AI respektującą wartości demokratyczne. Z kolei NIST rozwija dobrowolne mechanizmy zarządzania ryzykiem, kładąc nacisk na bezpieczeństwo już na etapie projektowania systemów. Te inicjatywy tworzą globalny język standardów, bez którego niemożliwe byłoby porozumienie ponad granicami. Technologia nie uznaje paszportów, toteż jej etyka musi posiadać charakter uniwersalny, wykraczający poza lokalne jurysdykcje.

Kluczowym punktem odniesienia dla globalnego bezpieczeństwa stał się AI Safety Report 2025, przygotowany pod redakcją Yoshuy Bengio. Raport ten stanowi chłodną syntezę ryzyk cywilizacyjnych, od dezinformacji i cyberataków aż po mroczne scenariusze dotyczące broni biologicznej czy utraty kontroli. Dokument nie jest kolejną marketingową broszurą, lecz rzetelną mapą zagrożeń, których nie wolno traktować jak literackiej fikcji czy marketingowej mgły. Perspektywa ta jest bliska idealizmowi, czyli filozoficznej ramie będącej odpowiedzią na

współczesne kryzysy prawdy i antropologii, choć zachowuje ona większą ostrożność normatywną. Zamiast obiecywać cyfrowe zbawienie, autorzy raportu wskazują na konkretne punkty zapalne, w których technologia może obrócić się przeciwko człowiekowi.

Aby zrozumieć fundamenty idealizmu i sens walki o podmiotowość, musimy powrócić do Kanta. Nie czynimy tego z akademickiego obowiązku, by każda myśl o moralności otrzymała pieczątkę z Królewca, lecz z konieczności znalezienia stałego lądu w płynnej rzeczywistości. Oświecenie jawi się tu jako projekt niedokończony, który w dobie serwerowni zyskuje zupełnie nową, dramatyczną aktualność. Jeśli bowiem niedojrzałość cyfrowa, czyli stan bierności, w którym rezygnujemy z krytycznego myślenia na rzecz gotowych odpowiedzi algorytmu, stanie się normą, stracimy szansę na autonomię. Algorytm nie poda się do dymisji, co najwyżej otrzyma patch, dlatego to my musimy rozstrzygnąć, czy technologia ma służyć człowiekowi, czy też stać się nowym, cyfrowym feudalizmem.

Sapere aude w epoce algorytmów: jak AI kształtuje naszą wolność

Kant pozostaje w tej debacie postacią absolutnie nieodzowną, ponieważ jego myśl spaja trzy fundamenty, które w dobie sztucznej inteligencji zyskują niemal dramatyczną aktualność: autonomię rozumu, godność osoby oraz projekt trwałego pokoju. Kantowskie Oświecenie nie było jedynie akademickim postulatem, lecz żarliwym wezwaniem do wyjścia z zawinionej niedojrzałości, rozumianej jako stan bierności intelektualnej. Owa niedojrzałość cyfrowa, czyli stan, w którym użytkownik rezygnuje z krytycznego myślenia na rzecz gotowych odpowiedzi generowanych przez algorytmy, nie wynika bynajmniej z braku inteligencji. Człowiek niedojrzały może wszak być wybitnie wykształcony, elokwentny i społecznie uprzywilejowany, a mimo to wciąż żyć cudzym sądem, nie znajdując w sobie odwagi, by posłużyć się własnym intelektem. W świecie przedcyfrowym kuratorami tej bierności były tradycyjne instytucje autorytetu, które narzucały dogmaty w sposób jawny i często toporny. Dziś jednak kuratela ta przybiera formę miękką, spersonalizowaną i niemal całkowicie niewidoczną dla nieuzbrojonego oka.

Współczesny system nie krzyczy już do nas „masz wierzyć”, lecz szepcze uwodzicielsko „to może cię zainteresować”, co czyni go znacznie skuteczniejszym narzędziem formatowania rzeczywistości. Zamiast zakazywać herezji, algorytmy po prostu serwują nam świat tak długo sortowany pod kątem naszych lęków i upodobań, aż sami zamieszkamy w epistemicznej klatce. Ściany tego więzienia wyglądają dokładnie tak, jak nasze własne preferencje, co sprawia, że rzadko kiedy odczuwamy potrzebę ucieczki. Tu właśnie ujawnia się potężna waga polityczna sztucznej inteligencji, która

operuje subtelniej niż jakikolwiek cenzor z minionej epoki. Potwór cyfrowy nie musi ryczeć; może mieć elegancki dashboard i model ryzyka, który po cichu zarządza naszą uwagą. Model rekomendacyjny nie musi bowiem niczego usuwać wprost, gdyż wystarczy mu jedynie umiejętne ustawienie kolejności widzialności poszczególnych treści.

W świecie informacyjnego nadmiaru władza nie polega już na zakazywaniu, lecz na zarządzaniu prawdopodobieństwem zauważenia konkretnych faktów przez użytkownika. Dawny despota palił książki na stosach, podczas gdy nowy porządek może je po prostu zignorować, czyniąc je niewidocznymi dla masowego odbiorcy. Jeśli algorytm uzna daną treść za słabo angażującą, to de facto przestaje ona istnieć w przestrzeni publicznej, mimo że formalnie nikt jej nie usunął. Jest to cenzura bez cenzora, miękka jak aksamit i skuteczna niczym precyzyjna księgowość, przed którą trudno się bronić tradycyjnymi metodami. W tym kontekście kantowskie „sapere aude" staje się w epoce AI imperatywem o charakterze czysto infrastrukturalnym, a nie tylko moralnym. Miej odwagę używać własnego rozumu, ale najpierw zadbaj o takie warunki techniczne, prawne i ekonomiczne, aby twój intelekt nie był od dzieciństwa formatowany przez systemy predykcyjnego bodźcowania.

Bez tej materialnej podstawy wezwanie do autonomii stanie się jedynie luksusowym sloganem dla nielicznych, których stać na wykupienie ciszy, edukacji i prywatności. Aidealizm, czyli filozoficzna rama będąca odpowiedzią na współczesne kryzysy, słusznie rozpoznaje, że Oświecenie nie jest zamkniętym rozdziałem w podręczniku historii. Jest to raczej nieustający projekt instytucjonalny, który wciąż bywa sabotowany przez ludzką naturę, głębokie nierówności oraz plemienne mitologie. Nowoczesność obiecała nam, że rozum publiczny zastąpi przemoc arbitralnej władzy i trzeba przyznać, że częściowo tę obietnicę spełniła poprzez konstytucjonalizm czy prawa człowieka. Niemniej jednak Oświecenie miało także swoje mroczne kontrapunkty, takie jak technokratyczny paternalizm czy złudzenie, że rozum instrumentalny sam z siebie stanie się rozumem moralnym. Właśnie w tym miejscu Aidealizm wykazuje swoją największą siłę, ponieważ nie postuluje on naiwnego powrotu do osiemnastowiecznych salonów.

Nie domaga się on „więcej racjonalności" w płaskim, menedżerskim sensie, lecz szuka raczej racjonalności refleksyjnej, solidarnej i antropologicznie świadomej. Potrzebujemy rozumu, który wie, że człowiek nie jest czystym podmiotem chłodnej kalkulacji, lecz istotą ucieleśnioną i podatną na manipulację. Sama informacja nie wyzwala, jeśli jej architektura uczy posłuszeństwa, co stanowi kluczową lekcję zarówno dla obywateli, jak i dla twórców modeli językowych. Algorytm może wszak przetworzyć całą bibliotekę świata, ale to wcale nie oznacza, że automatycznie stworzy on republikę rozumu. Prawdziwa republika wymaga bowiem nie tylko dostępu do danych, lecz przede wszystkim procedur sporządzania sądu oraz kultury odpowiedzialnego sporu. Aidealizm trafnie podkreśla

przy tym paradoks ludzkiej natury, traktując ją jako swoiste ryzyko systemowe, którego nie wolno nam ignorować.

Jesteśmy zdolni do wielkiej empatii, ale potrafimy też rano napisać deklarację praw człowieka, by po południu znaleźć wygodny wyjątek dla naszego wroga. W naukach społecznych ta diagnoza znajduje liczne potwierdzenia, wskazując na nasze ograniczenia poznawcze i skłonność do moralnego upiększania instynktów. Psychologia poznawcza od dekad dowodzi, że ludzkie decyzje są więźniami heurystyk i motywowanego rozumowania, co czyni nas łatwym celem dla algorytmicznej perswazji. Ekonomia behawioralna podważyła z kolei obraz aktora racjonalnego, pokazując, że reagujemy silniej na ramowanie i status quo niż na obiektywną użyteczność. Antropologia polityczna przypomina nam natomiast, że wspólnoty budują swoją tożsamość poprzez narracje o zagrożeniu, co algorytmy potrafią wykorzystać z zabójczą precyzją. Socjologia wiedzy dodaje do tego istotne zastrzeżenie: fakty społeczne żyją w instytucjach zaufania, a gdy one upadają, sama prawda przestaje wystarczać. Barbarzyństwo z interfejsem premium nadal pozostaje barbarzyństwem, nawet jeśli ubrane jest w elegancki raport zgodności i nowoczesny design.

Między boską iskrą a algorytmem: granice ludzkiej twórczości

AI wchodzi w zastane struktury społeczne nie jako neutralne narzędzie, lecz jako potężny akcelerator naszych dotychczasowych tendencji. Może ona pomagać w korygowaniu błędów poznawczych, ale równie dobrze potrafi dostarczać każdej grupie ideologicznej jej własne, syntetyczne potwierdzenia. Potrafi wykrywać manipulacje, lecz jednocześnie umożliwia produkcję milionów mikrowpływów dostosowanych do profilu psychologicznego konkretnego odbiorcy. Może zwiększyć dostęp do edukacji, ale może też utrwalić stan intelektualnej bierności, jeśli użytkownik przestanie pytać, a zacznie jedynie bezrefleksyjnie konsumować odpowiedzi. Problemem nie jest zatem to, że AI okaże się „nieludzka", lecz to, że może być zbyt ludzka w najgorszym sensie. Została przecież wytrenowana na naszych uprzedzeniach, reklamach, sporach, fantazjach i aktach przemocy symbolicznej.

Klasyczna maksyma Nietzschego o walce z potworami brzmi w epoce algorytmów jak ostrzeżenie przed niebezpiecznym sprzężeniem zwrotnym. Projektujemy systemy do walki z chaosem informacyjnym, lecz jeśli nakarmimy je logiką dominacji, same nieuchronnie staną się proceduralnym potworem. Potwór cyfrowy nie musi ryczeć; może mieć dashboard, model ryzyka i elegancki raport zgodności, co czyni go niemal niewidocznym. Dlatego idealizm AI, czyli filozoficzna rama będąca odpowiedzią na współczesne kryzysy, słusznie postuluje, by AI stała się

protezą dojrzałości, a nie lenistwa. Ma ona wzmacniać naszą zdolność do samodzielnego myślenia, zamiast zachęcać do całkowitego outsourcingu procesów poznawczych. Prawdziwy postęp polega na wspieraniu rozumu publicznego, a nie na zastępowaniu go automatycznym autorytetem kodu.

Kantowskie „wyjście z niedojrzałości" w epoce cyfrowej oznacza nie tylko edukację obywateli, lecz także audyt instytucji i kontrolę ekonomii uwagi. Niedojrzałość cyfrowa, czyli stan bierności intelektualnej, w którym rezygnujemy z krytycyzmu na rzecz gotowych algorytmicznych odpowiedzi, zagraża naszej autonomii. Jeśli nie nauczymy się zarządzać tymi systemami świadomie, będziemy mieć społeczeństwo pozornie poinformowane, lecz realnie prowadzone na smyczy bodźców. Musimy zatem zapytać, co właściwie rozumiemy przez twórczość w świecie zdominowanym przez czyste obliczenia. Czy potrafimy jeszcze odróżnić boską iskrę od sprawnego, lecz ostatecznie pustego w środku algorytmicznego pastiszu? To pytanie staje się punktem wyjścia dla nowej antropologii, która musi na nowo zdefiniować granice między człowiekiem a maszyną.

Jednym z najbardziej prowokacyjnych wątków współczesnej debaty jest rozróżnienie między ludzką kreatywnością de novo a obliczeniowym pastiszem generowanym przez modele. Warto bronić tej granicy, zachowując jednak naukową ostrożność i unikając łatwych stwierdzeń, że maszyna „nigdy" nie stworzy czegoś nowego. Historia technologii jest przecież cmentarzem pewnych siebie niemożliwości, które z czasem stały się naszą codziennością. Jeśli kreatywność definiujemy jedynie jako generowanie statystycznie nowej kombinacji elementów, to AI bez wątpienia jest twórcza już w tej chwili. Potrafi ona komponować obrazy, teksty i melodie, które nie istniały wcześniej w dokładnie takiej postaci. Jeśli jednak kreatywność rozumiemy jako akt egzystencjalny, sytuacja ulega radykalnej zmianie.

Prawdziwa twórczość wyrasta z doświadczenia świata, cielesności, pragnienia, lęku, pamięci, straty, winy oraz poczucia odpowiedzialności. Maszyna może symulować formę lamentu, ale nie utraciła nikogo, kogo mogłaby opłakiwać w swoim cyfrowym wnętrzu. Może napisać modlitwę, ale nie potrafi klęknąć, ponieważ nie zna ciężaru sacrum ani lęku przed ostatecznością. Może wygenerować obraz samotności, ale nie zna nocy, w której samotność przestaje być tematem, a staje się temperaturą pokoju. To nie jest sentymentalny argument przeciw technologii, lecz twarda teza antropologiczna o zakorzenieniu sztuki w śmiertelności. Ludzka twórczość wynika z faktu, że człowiek nie tylko przetwarza znaki, lecz stawia na szali własną osobę.

Twórca ryzykuje kompromitację, odsłonięcie, porażkę i odrzucenie, a niekiedy nawet utratę własnej tożsamości w procesie kreacji. Model generatywny nie ryzykuje niczego, nie ma bowiem honoru, który można zranić, ani życia, którego nie da się powtórzyć. Krytycy tacy jak Emily Bender słusznie wskazują, że wielkie modele językowe to „stochastyczne papugi", generujące wypowiedzi bez rozumienia sensu. To ważna korekta wobec taniego zachwytu nad technologiczną elegancją, która

często maskuje brak jakiegokolwiek sumienia. Model może mówić jak mędrzec, nie będąc mędrcem, i imitować styl odpowiedzialności, nie mogąc za nic odpowiedzieć. Barbarzyństwo z interfejsem premium nadal pozostaje barbarzyństwem, nawet jeśli posługuje się nienaganną składnią.

Idealizm AI nie powinien jednak wpadać w skrajność i traktować sztucznej inteligencji wyłącznie jako maszynki do kulturowego plagiatu. Technologia ta może być realnym katalizatorem ludzkiej kreatywności, poszerzając pole skojarzeń i przyspieszając procesy iteracji. Może stać się muzą, pod warunkiem jednak, że ta muza nie będzie próbowała podpisać aktu własności duszy artysty. Granica jest jasna: AI może rozszerzać ludzką inwencję, ale nie powinna kolonizować naszej sprawczości, czyli zdolności do świadomego działania. W praktyce oznacza to konieczność nowej etyki autorstwa, jawności użycia systemów oraz ochrony realnej pracy twórczej. Musimy zadbać o to, by edukacja nie myliła sprawnego promptowania z głębokim, samodzielnym myśleniem.

Algorytmiczny pastisz jest niebezpieczny nie dlatego, że jest brzydki, lecz dlatego, że może obniżyć próg kulturowej ambicji. Społeczeństwo karmione wyłącznie syntetyczną przyjemnością estetyczną ryzykuje utratę smaku trudności, bez którego nie ma wysokiej kultury. Pozostaje wtedy jedynie premium content dla zmęczonych oczu, pozbawiony głębi i autentycznego wyzwania intelektualnego. Uczeń zapytał maszynę, czym jest geniusz, a ta odpowiedziała natychmiast, cytując definicje i modele psychometryczne. Kiedy jednak zapytał starego kompozytora, ten milczał długo, po czym powiedział, że geniusz to ból błędu uczący słyszeć ciszę. Maszyna zapisała tę odpowiedź do korpusu, lecz uczeń zapamiętał ją w ciele, doświadczając przemiany, której kod nie zna.

Dlaczego Aidealizm jest niezbędną odpowiedzią na kryzysy AI

Aidealizm nie jest estetycznym dodatkiem do technologicznego dyskursu, lecz surową koniecznością wynikającą z faktu, że stoimy obecnie w obliczu pięciu nakładających się na siebie kryzysów. Musimy zmierzyć się z erozją prawdy, destabilizacją pracy, uwiądem demokracji, zapaścią ekologiczną oraz fundamentalnym kryzysem antropologicznym, często objawiającym się jako niedojrzałość cyfrowa, czyli stan bierności intelektualnej, w którym użytkownik rezygnuje z krytycznego myślenia na rzecz gotowych odpowiedzi generowanych przez algorytmy. Jeśli nie stworzymy silnej ramy filozoficznej, sztuczna inteligencja zostanie błyskawicznie wchłonięta przez najsilniejsze, a przy tym najmniej niewinne logiki współczesności. Logika platformowa pożąda naszej uwagi, finansowa żąda zwrotu z kapitału, a autorytarna marzy o pełnej przewidywalności obywatela. Do tego dochodzi logika wojskowa szukająca przewagi, biurokratyczna dążąca do

standaryzacji oraz marketingowa, która chce totalnego wpływu na nasze zachowania. Bez normatywnej kontrlogiki otrzymamy system, który z matematyczną precyzją będzie formatował i monetyzował ludzkie życie. Barbarzyństwo z interfejsem premium nadal pozostaje barbarzyństwem, nawet jeśli jego algorytmy są trenowane na najdroższych procesorach świata.

Aidealizm, czyli filozoficzna rama i odpowiedź na współczesne kryzysy, przesuwa ciężar dowodu na tych, którzy wierzą w zbawczą moc niewidzialnej ręki rynku w świecie kodu. Jeśli ktoś twierdzi, że rozwój AI powinien toczyć się wyłącznie według reguł komercyjnych, musi najpierw wyjaśnić, dlaczego ten sam rynek zawiódł nas w niemal każdej innej cyfrowej domenie. Nie potrafił on wszak samodzielnie ochronić naszej prywatności, powstrzymać monopolizacji platform ani zabezpieczyć jakości debaty publicznej przed zalewem dezinformacji. Wiara w to, że drapieżny oligopol nagle wytworzy etyczną superstrukturę, nie jest liberalizmem, lecz czystą metafizyką zarządu, która ignoruje historyczne lekcje o naturze władzy. Aidealizm domaga się wpisania technologii w projekt godności i solidarności, ponieważ technologia nigdy nie rozwija się w aksjologicznej próżni. Wiara w to, że algorytmiczny gigant sam z siebie narzuci sobie kaganiec, przypomina ufność w to, że grawitacja przestanie działać na prośbę dewelopera.

Należy jednak zachować czujność, gdyż Aidealizm nie jest prostym etatyzmem, a państwo bywa drapieżnikiem równie skutecznym co wielkie korporacje technologiczne. Państwo dysponujące potęgą obliczeniową może łatwo porzucić klasyczną administrację na rzecz predykcyjnego zarządzania populacją, przy czym może naruszać naszą nieprzejrzystość egzystencjalną, czyli prawo jednostki do posiadania sfery życia, która nie jest stale interpretowana i punktowana przez systemy obliczeniowe. Rozpoznawanie twarzy, scoring obywateli oraz automatyczne profilowanie mogą razem stworzyć panoptykon, o jakim dawni despoci mogli tylko marzyć w chwilach największej pychy. Toteż właśnie potrzebujemy trzech wzajemnie ograniczających się sił: demokratycznego państwa prawa, pluralistycznego społeczeństwa obywatelskiego oraz kontrolowanej innowacji gospodarczej. Żadna z tych sfer nie może otrzymać monopolu na rozwój AI, gdyż każda z nich, pozostawiona bez nadzoru, dąży do totalności. Równowaga ta jest niezbędna, aby technologia służyła wolności, a nie stawała się narzędziem cyfrowego feudalizmu.

Współczesna krytyka dostarcza Aidealizmowi fundamentów, odzierając AI z jej rzekomej, niematerialnej aury. Kate Crawford przypomina, że chmura to w rzeczywistości cudzy komputer, kopalnia litu, ogromne ilości wody i często niewidzialna, nisko płatna praca. Z kolei badacze tacy jak Arvind Narayanan ostrzegają przed „AI snake oil", czyli myleniem realnych możliwości z predykcyjną alchemią, która obiecuje więcej, niż matematyka może dowieść. Taka epistemiczna higiena jest kluczowa, by Aidealizm nie zmienił się w technoutopijną liturgię, oderwaną od fizycznych ograniczeń świata. Musimy odróżniać ziarno postępu od plew marketingowego szumu,

który karmi się naszymi lękami i nadziejami. Tylko trzeźwe spojrzenie na infrastrukturę materialną pozwala zrozumieć prawdziwą cenę, jaką płacimy za cyfrowe udogodnienia.

W debacie publicznej ścierają się różne wizje przyszłości, od alarmizmu egzystencjalnego po sceptycyzm Yanna LeCuna, który kwestionuje nieuchronność zagrożenia ze strony AGI. Choć spór o bunt maszyn jest fascynujący, Aidealizm skupia się przede wszystkim na tym, co dzieje się tu i teraz. Nie musimy czekać na apokalipsę, skoro cyfrowy feudalizm i automatyzacja nadzoru już teraz siedzą w naszym salonie i szukają gniazdka do ładowania. Ryzyka społeczne i polityczne są realne w tej sekundzie, objawiając się w erozji pracy, deformacji wiedzy oraz naruszaniu ludzkiej godności. Wystarczy wziąć pod uwagę te aktualne zagrożenia, aby uzasadnić potrzebę nowego kontraktu społecznego, który ureguluje nasze relacje z algorytmami. Aidealizm jest odpowiedzią na konkretne niesprawiedliwości, a nie tylko teoretycznym ćwiczeniem z zakresu futurologii.

Ostatecznie chodzi o to, by AI stała się narzędziem emancypacji, a nie kolejnym mechanizmem wykluczenia, co wymaga wprowadzenia konkretnych rozwiązań systemowych. Jednym z nich jest powszechna dywidenda AI, czyli koncepcja sprawiedliwego podziału korzyści ekonomicznych płynących z automatyzacji, która ma zapobiegać drastycznym nierównościom. Bez takich mechanizmów grozi nam nierówność antropotechniczna, czyli podział ludzkości wynikający z nierównego dostępu do sprywatyzowanych ulepszeń biologicznych i poznawczych. Musimy zatem dążyć do tego, by owoce postępu nie zostały przejęte przez nielicznych, lecz służyły całemu społeczeństwu w sposób transparentny i rozliczalny. Cywilizacja nie ginie bowiem dlatego, że zabrakło jej narzędzi, lecz dlatego, że narzędzia przeżyły jej sumienie.

Sześć filarów Aidealizmu: architektura dla cywilizacji AI

Rama AIDEAL, czyli filozoficzna struktura będąca odpowiedzią na kryzysy prawdy i demokracji, nie powinna być postrzegana jako zbiór haseł, lecz jako architektura instytucjonalna. Jej pierwszy filar, odpowiedzialne zarządzanie, mierzy się z pytaniem: kto odpowiada za system działający autonomicznie i na wielką skalę? W tradycyjnym prawie odpowiedzialność przypisywano konkretnej osobie lub producentowi, lecz AI skutecznie rozmywa te klasyczne kategorie. Decyzja algorytmiczna jest wszak wynikiem splotu danych, modelu i kontekstu, co tworzy niebezpieczną lukę w systemie sprawiedliwości. Jeśli prawo nie wypracuje czytelnych reżimów odpowiedzialności, zaakceptujemy najwygodniejszą fikcję naszych czasów: winny będzie algorytm, czyli nikt. Algorytm nie poda się przecież do dymisji, co najwyżej otrzyma patch.

Drugi filar, integralność intelektualna, stanowi odpowiedź na kryzys epistemologii publicznej w środowisku nasyconym treściami syntetycznymi. Tutaj AI powinna pełnić rolę narzędzia

autentykacji i pluralizacji poznawczej, zamiast stawać się źródłem wszechobecnej dezinformacji. Nie dążymy do powołania centralnego arbitra prawdy, gdyż byłaby to prosta droga do budowy ministerstwa epistemicznego paternalizmu. Niemniej chodzi raczej o warunki, w których obywatel zachowuje zdolność do weryfikacji twierdzeń i odróżnienia błędu od celowej propagandy. Musimy unikać stanu niedojrzałości cyfrowej, czyli bierności intelektualnej, w której rezygnujemy z krytycznego myślenia na rzecz gotowych odpowiedzi. Tylko w ten sposób utrzymamy wspólny świat faktów, będący niezbędnym fundamentem każdej demokracji.

Trzeci filar, rozproszony dobrobyt, jest ekonomicznym sercem Aidealizmu i rzuca wyzwanie logice wąskiej prywatyzacji zysków. AI buduje swoją wartość na fundamencie danych społecznych, publicznej infrastruktury oraz ogromnych nakładów energii. Jeśli owoce tej pracy zostaną przejęte przez nielicznych, a koszty uspołecznione, powstanie mechanizm oligarchiczny o nowej sile. Cyfrowy feudalizm nie oznacza powrotu do zamków, lecz sytuację, w której dostęp do mocy obliczeniowej staje się nową formą prawa senioralnego. Toteż w tym kontekście niezbędna jest powszechna dywidenda AI, czyli podział korzyści płynących z automatyzacji, zapobiegający drastycznym nierównościom. Bez niej inteligencja operacyjna gospodarki stanie się narzędziem arystokracji, która wjechała windą na szczyt i ogłosiła, że schody są dla słabych.

Czwarty filar, etyczna innowacja, chroni nas przed kultem skuteczności, który często przesłania sens działań. Innowacja nie staje się moralnie dobra tylko dlatego, że jest nowa, bowiem można przecież innowacyjnie manipulować lub nadzorować społeczeństwo. Etyczna innowacja pyta zatem o kierunek rozwoju, a nie tylko o jego tempo, analizując wpływ technologii na autonomię jednostki. Musimy rozstrzygnąć, czy dane rozwiązanie poszerza dostęp do dobra, czy może produkuje jedynie kulturę algorytmicznie strawionej przeciętności. Barbarzyństwo z interfejsem premium nadal pozostaje barbarzyństwem, nawet jeśli ukryte jest pod płaszczem zaawansowanego kodu.

Piąty filar, zaawansowane zrównoważenie, osadza rozwój technologii w realiach granic planetarnych, których nie da się zignorować. Jest to postulat konieczny, gdyż nie istnieje etyczna AI funkcjonująca na martwej planecie. Choć modele mogą wspierać transformację energetyczną, same generują ogromne koszty w postaci zużycia wody, rzadkich materiałów oraz energii. Każda odpowiedzialna polityka musi zatem liczyć ten rachunek, uwzględniając również koszty geopolityczne. W przeciwnym razie będziemy rozwiązywać kryzys klimatyczny przy pomocy infrastruktury, która po cichu dopisuje własny rachunek do atmosfery. Potwór cyfrowy nie musi wszak ryczeć, może mieć elegancki dashboard i raport zgodności.

Szósty filar, wolność i godność człowieka, stanowi klamrę spinającą całość i nadającą architekturze sens. Bez tego fundamentu pozostałe zasady mogłyby łatwo stać się sprawną, lecz pozbawioną ducha technokratyczną maszynerią. Musimy chronić nieprzejrzystość egzystencjalną, czyli prawo

do sfery życia, która nie jest stale interpretowana i punktowana przez systemy obliczeniowe. Cywilizacja nie ginie najczęściej dlatego, że zabrakło jej narzędzi, lecz dlatego, że narzędzia te przeżyły jej sumienie. Naszym zadaniem jest budowa systemu, w którym technologia służy człowiekowi, a nie staje się narzędziem jego cyfrowego zniewolenia. Tylko poprzez te zasady unikniemy przyszłości, w której człowiek staje się jedynie dodatkiem do własnych wynalazków.

Ekonomia jako próba ognia dla etyki sztucznej inteligencji

Godność to w istocie fundamentalna deklaracja, że człowiek nie jest jedynie sumą cyfrowych śladów ani funkcją w algorytmie przewidywania zachowań, lecz bytem autonomicznym. Wolność z kolei zakłada, że nie każda skuteczna personalizacja, choćby najbardziej kusząca, jest etycznie dopuszczalna w demokratycznym społeczeństwie. Kluczowa staje się tutaj nieprzejrzystość egzystencjalna, czyli prawo jednostki do posiadania sfery życia, która nie jest stale interpretowana, punktowana i przewidywana przez systemy obliczeniowe. Człowiek nie musi być przecież całkowicie czytelny dla systemu, aby zachować status dobrego obywatela. Państwo i rynek, żądając pełnej przejrzystości, rzadko dążą do autentycznego zrozumienia kondycji ludzkiej. Chcą one po prostu sprawniej nami zarządzać.

W tym kontekście aidealizm, czyli filozoficzna rama będąca odpowiedzią na współczesne kryzysy prawdy i antropologii, jawi się jako próba ocalenia dziedzictwa Oświecenia w epoce zautomatyzowanego rozumu. Nie jest to projekt naiwny, o ile rozumiemy go jako rygorystyczną filozofię kontroli nad potężną mocą obliczeniową, a nie kolejny hymn pochwalny na cześć krzemowych bóstw. Aidealizm broni tezy, że sztuczna inteligencja musi pozostać podporządkowana dobru wspólnemu, prawdzie i solidarności. Pozostawiona wyłącznie logice zysku lub militarnej dominacji, stanie się jedynie precyzyjnym narzędziem deformacji demokracji. Potwór cyfrowy nie musi przecież ryczeć; może mieć elegancki pulpit nawigacyjny, model ryzyka i nienaganny raport zgodności.

Kantowskie zaplecze tych rozważań, oparte na naturze ludzkiej i różnicy między kreatywnością a czystym obliczeniem, musi jednak zejść z wyżyn metafizyki do miejsca znacznie mniej wzniosłego, lecz rozstrzygającego: do ekonomii. Historia uczy nas bowiem brutalnej lekcji: wielkie ideały rzadko kruszą się w ogniu debat filozoficznych, znacznie częściej dogorywają w arkuszach kalkulacyjnych. Wolność bywa wpisana do konstytucji, by ostatecznie polec w starciu z modelem biznesowym. Godność, ogłoszona prawem niezbywalnym, w procesie optymalizacji zbyt łatwo zamienia się w zwykłą zmienną kosztową. Prawda zaś, choć czczona w akademiach, bywa bezlitośnie sortowana

według wskaźnika zaangażowania użytkowników. Jeśli więc pytamy o przyszłość AI, musimy zapytać o to, kto ją posiada i kto czerpie z niej rentę.

Ekonomia sztucznej inteligencji nie jest zatem jedynie przypisem do filozofii, lecz jej ostateczną próbą ognia, w której hartują się nasze deklaracje. Można wygłaszać najpiękniejsze manifesty o humanizmie technologicznym, ale jeśli infrastruktura, talenty i kapitał zostaną skupione w rękach nielicznych graczy, język etyki stanie się tylko ceremonialnym kadzidłem nad nowym feudalizmem. Aidealizm postuluje, by technologia ta służyła rozproszonemu dobrobytowi, zapobiegając koncentracji władzy ekonomicznej na niespotykaną dotąd skalę. Współczesne dane każą traktować to ostrzeżenie z najwyższą powagą, a nie jako literacką figurę retoryczną.

Twarde dane ze Stanford AI Index 2025 pokazują, że nie mamy już do czynienia z fazą niewinnych eksperymentów, lecz z brutalnym wyścigiem zbrojeń finansowych. Prywatne inwestycje w AI w Stanach Zjednoczonych osiągnęły w 2024 roku poziom 109,1 miliarda dolarów, co czyni je niemal dwunastokrotnie większymi niż w Chinach. Globalnie sama generatywna AI przyciągnęła blisko 34 miliardy dolarów, a odsetek organizacji deklarujących jej wykorzystanie skoczył z 55 do 78 procent w ciągu zaledwie jednego roku. W świecie, gdzie algorytm nie poda się do dymisji, a co najwyżej otrzyma poprawkę, te liczby definiują nową architekturę władzy. To nie jest już tylko technologia; to fundament nowego porządku świata.

Cyfrowy feudalizm i sprawiedliwa dywidenda z AI

Obecnie obserwujemy fazę kolonizacji infrastruktury gospodarczej przez inteligencję maszynową, co stanowi proces głęboko przekształcający fundamenty naszej wymiany rynkowej. Sygnał ostrzegawczy płynie z raportu UNCTAD Technology and Innovation Report 2025, który wskazuje na niebezpieczną koncentrację rozwoju AI w rękach nielicznych korporacji. Takie luki infrastrukturalne mogą drastycznie powiększać nierówności, tworząc przepaść między cyfrowymi metropoliami a peryferiami, które pozostają jedynie biernymi odbiorcami technologii. Choć AI obiecuje bezprecedensowy wzrost produktywności, nie istnieje żaden obiektywny mechanizm rynkowy, który gwarantowałby sprawiedliwy podział wypracowanych w ten sposób owoców. Jeżeli fundamenty systemu pozostają w posiadaniu nielicznych, to korzyści płynące z ich eksploatacji również zostaną skumulowane na samym szczycie hierarchii. To stara lekcja ekonomii politycznej ubrana w nową, krzemową szatę, przypominająca nam, że technologia bez etyki staje się narzędziem wykluczenia.

Pojęcie cyfrowego feudalizmu bywa nadużywane w publicystyce, jednak w kontekście rozwoju sztucznej inteligencji nabiera ono bardzo precyzyjnego i niepokojącego sensu. Nie chodzi tu o

malowniczy powrót do epoki rycerzy i lenn, choć współcześni technobaroni zapewne chętnie przyjęliby taką estetykę dla podkreślenia swojej dominacji. Feudalizm oznacza tutaj specyficzną strukturę zależności, w której większość uczestników życia gospodarczego korzysta z infrastruktury, której w żaden sposób nie posiada. Produkujemy dane, nad którymi nie mamy kontroli, podlegamy algorytmicznym regułom, których nie negocjujemy, i płacimy rentę podmiotom zarządzającym warunkami dostępu. Potwór cyfrowy nie musi ryczeć, może mieć dashboard, model ryzyka i elegancki raport zgodności, a mimo to skutecznie drenować naszą autonomię. W tej nowej architekturze władzy wolność wyboru staje się jedynie starannie zaprojektowaną iluzją wewnątrz zamkniętego ekosystemu.

W klasycznym kapitalizmie przemysłowym robotnik sprzedawał swoją pracę, natomiast w kapitalizmie platformowym użytkownik nieświadomie handluje własną uwagą i czasem. Kapitalizm AI idzie o krok dalej, czyniąc z całego społeczeństwa dostawcę surowca poznawczego, źródło wzorców behawioralnych oraz przedmiot nieustannych predykcji. Stajemy się rynkiem zbytu dla systemów, które następnie zautomatyzują znaczną część naszych kompetencji, co tworzy wyjątkowo osobliwy i asymetryczny kontrakt społeczny. Oddajemy najcenniejszy surowiec, jakim jest nasze doświadczenie, a następnie kupujemy dostęp do maszyn, które dzięki temu surowcowi zwiększają przewagę właścicieli infrastruktury. Barbarzyństwo z interfejsem premium nadal pozostaje barbarzyństwem, nawet jeśli jest opakowane w obietnicę optymalizacji każdego aspektu naszej codziennej egzystencji.

Idealizm, czyli filozoficzna rama będąca odpowiedzią na współczesne kryzysy, celnie zauważa, że pytanie o własność danych nie jest jedynie technicznym problemem dotyczącym prywatności. Jest to fundamentalne pytanie o sprawiedliwość dystrybucyjną, ponieważ dane nie są jedynie śladem jednostkowym, lecz stanowią dobro o charakterze relacyjnym i społecznym. Mój wynik medyczny zyskuje realną wartość dopiero wtedy, gdy można go porównać z wynikami tysięcy innych pacjentów w ramach szerszej populacji. Mój styl pisania ma wszak znaczenie tylko dlatego, że istnieje ogromny korpus stylów, na tle których algorytm może wyodrębnić specyficzne cechy mojej ekspresji. Jeśli więc sztuczna inteligencja karmi się zbiorowym życiem społecznym, nie można całej wartości wygenerowanej z tego procesu przypisać wyłącznie właścicielowi serwera.

Tutaj rodzi się argument na rzecz powszechnej dywidendy AI, czyli koncepcji sprawiedliwego podziału korzyści ekonomicznych płynących z automatyzacji i wzrostu produktywności maszynowej. Nie chodzi tu o jałmużnę dla ludzi wypchniętych z rynku pracy, lecz o uznanie, że społeczeństwo jest pełnoprawnym współproducentem wartości algorytmicznej. Dane, język, instytucje oraz edukacja publiczna stanowią część ukrytego kapitału, na którym budowane są współczesne potęgi technologiczne. Toteż powszechna dywidenda AI nie jest prezentem od

technologicznych możnych, lecz formą czynszu cywilizacyjnego należnego wspólnocie za używanie jej poznawczego podłoża. Stabilność prawna oraz kultura zaufania są fundamentem, bez którego żadna innowacja nie mogłaby przynieść realnego zysku. Sprawiedliwość w erze algorytmów wymaga przedefiniowania pojęcia własności i uznania wkładu każdego z nas w ten wspólny projekt.

Ekonomia Aidealizmu: redystrybucja zysków i własność danych

Ten argument zyskuje dodatkową wagę w świetle chłodnych prognoz dotyczących globalnego rynku pracy, które rzucają cień na niemal 40 procent dotychczasowych stanowisk. Międzynarodowy Fundusz Walutowy wskazuje, że w gospodarkach zaawansowanych ekspozycja na zmiany jest jeszcze wyższa, przy czym nie mamy tu do czynienia z prostym scenariuszem technologicznego bezrobocia. Jest to raczej subtelna gra, w której część ról zostanie przez algorytmy uzupełniona, podczas gdy inne zostaną bezpowrotnie zastąpione przez cyfrowe odpowiedniki. AI nie jest bowiem jedynie technologiczną nowinką, lecz nowym aktorem w odwiecznym dramacie redystrybucji ryzyka, dochodu oraz siły przetargowej. Jeśli nowa technologia zwiększy produktywność wyłącznie wysoko wykwalifikowanych pracowników i właścicieli kapitału, otrzymamy bolesną wersję polaryzacji rynku. Nie będzie ona jednak przypominać dawnej, industrialnej walki fabryki z robotnikiem, którą znamy z podręczników historii.

Dzisiejszy konflikt jest znacznie bardziej wyrafinowany, niemal aksamitny w swojej bezwzględności, gdyż uderza w same fundamenty eksperckiej wiedzy. Wyobraźmy sobie specjalistę, który zostaje „wsparty” przez systemy tak intensywnie, że jego unikalne dotąd kompetencje stają się nagle zasobem całkowicie wymiennym. Analityk otrzymuje do rąk narzędzia, które początkowo przyspieszają jego pracę, by w kolejnym kroku drastycznie zmniejszyć zapotrzebowanie na kolejnych pracowników w jego dziale. Ostatecznie system przedefiniuje standard produktywności w taki sposób, że dawny, stabilny etat zacznie wyglądać jak luksusowa nieefektywność z minionej epoki. To nie jest bynajmniej antytechnologiczny lament, lecz trzeźwe pytanie o to, kto w rzeczywistości przejmie nadwyżkę płynącą z powszechnej automatyzacji. Algorytm nie poda się przecież do dymisji, co najwyżej otrzyma aktualizację, podczas gdy człowiek zostanie z rachunkiem za swoją nagłą zbędność.

W klasycznej ekonomii wzrost produktywności miał być fundamentem wzrostu płac, niemniej w gospodarce opartej na AI ten związek może zostać zerwany brutalniej niż kiedykolwiek wcześniej. Jeśli marginalny koszt reprodukcji pracy poznawczej spadnie niemal do zera, a potężne modele pozostaną własnością nielicznych, wynagrodzenia stracą związek z wartością społeczną zadań.

Powstanie wtedy swoisty paradoks: nasze społeczeństwo będzie teoretycznie bogatsze w możliwości, lecz miliony ludzi staną się dramatycznie biedniejsze w sprawczość. Będziemy świadkami nadprodukcji treści, analiz i optymalizacji, przy jednoczesnym zaniku podmiotowego udziału jednostki w procesie tworzenia realnej wartości. Cywilizacja w takim wydaniu zacznie przypominać luksusowy hotel, w którym większość ludzkości, mimo eleganckiej fasady, zamieszkuje w dusznej piwnicy technicznej. Aidealizm, czyli filozoficzna rama i odpowiedź na współczesne kryzysy prawdy, pracy, demokracji, ekologii oraz antropologii, podpowiada nam tutaj gorzki bon mot: nie ma postępu, jeśli zyski są sztuczne, a koszty pozostają boleśnie ludzkie.

W epoce przemysłowej fundamentem państwa opiekuńczego było opodatkowanie pracy, przedsiębiorstw oraz konsumpcji, co pozwalało na utrzymanie elementarnej spójności społecznej. Jeśli jednak sztuczna inteligencja drastycznie zmniejszy udział ludzkiego wysiłku w produkcji wartości, klasyczny system fiskalny zostanie wystawiony na próbę, której może nie przetrwać. Automatyzacja potrafi wprawdzie windować PKB i zyski korporacyjne, lecz jednocześnie skutecznie ogranicza wpływy powiązane z zatrudnieniem, podcinając finansowanie edukacji czy ochrony zdrowia. Państwo, które nie zdoła w porę przemyśleć tej fundamentalnej zmiany, ryzykuje, że stanie się jedynie fiskalnym muzeum minionej epoki etatu. Dlatego konieczne stają się podatki od produktywności AI, od ekstrakcji danych oraz od algorytmicznych transakcji finansowych, choć każdy z nich wymaga chirurgicznej precyzji.

Źle zaprojektowana danina może wprawdzie zniechęcić do innowacji lub wypchnąć kapitał do jurysdykcji o niższych wymogach, niemniej brak opodatkowania oznacza znacznie groźniejszą formę arbitrażu. Mamy tu bowiem do czynienia ze starym numerem w nowym, cyfrowym garniturze, czyli prywatyzacją zysków przy jednoczesnym uspołecznieniu wszelkich kosztów transformacji. Kiedy model generuje dochód, trafia on do prywatnej kieszeni, lecz gdy ten sam system dyskryminuje użytkowników lub destabilizuje rynek pracy, rachunek płaci społeczeństwo. Podatek od produktywności AI nie powinien być zatem rozumiany jako kara za postęp, lecz jako niezbędny mechanizm udziału obywateli w wartości powstałej dzięki maszynom. Podatek od danych powinien natomiast wyrażać fakt, że informacje nie są darmowym darem natury, lecz rezultatem naszego wspólnego życia, relacji i zachowań.

Opodatkowanie transakcji algorytmicznych mogłoby z kolei ograniczyć skrajnie spekulacyjne działania systemów wysokiej częstotliwości, które budują przewagę gigantów, nie służąc przy tym realnej gospodarce. Nie miejmy jednak złudzeń, że budowa nowej architektury fiskalnej będzie zadaniem łatwym, skoro wartość generowana przez modele jest tak trudna do terytorialnej alokacji. Globalne korporacje to wszak mistrzowie optymalizacji, a sztuczna inteligencja dostarcza im jedynie nowych, wyrafinowanych warstw do tej cynicznej gry. Dlatego też Aidealizm musi zyskać

wymiar międzynarodowy, gdyż powszechna dywidenda AI, czyli koncepcja sprawiedliwego podziału korzyści ekonomicznych płynących z automatyzacji i wzrostu produktywności maszynowej, w jednym kraju pozostanie jedynie kruchym eksperymentem. Dopiero globalna koordynacja podatkowa mogłaby stać się początkiem prawdziwej sprawiedliwości algorytmicznej, która uwzględniałaby również palący problem energetyczny.

Sztuczna inteligencja nie bytuje przecież w metafizycznym eterze; jej „chmura" jest materialna, przerażająco energochłonna i twardo osadzona w konkretnej geografii. Dane Międzynarodowej Agencji Energii są tu bezlitosne: w 2025 roku globalne zapotrzebowanie centrów danych na energię wzrosło o 17 procent, a w przypadku systemów AI ten wzrost sięgnął aż 50 procent. Takie liczby skutecznie niszczą bajkę o niematerialnej inteligencji, która rozwija się bez wystawiania rachunku ekologicznego naszej planecie. Aidealizm domaga się więc, by każda poważna debata o ekonomii AI obejmowała rzetelny rachunek zużycia wody, energii, rzadkich minerałów oraz infrastruktury przesyłowej. AI może wprawdzie wspierać optymalizację sieci energetycznych, niemniej napędzana logiką nieograniczonego wzrostu obliczeń, szybko stanie się częścią problemu, a nie jego rozwiązaniem.

Prawdziwy postęp nie polega przecież na tym, że dymiąca spalarnia zostaje eufemistycznie nazwana „centrum innowacji" w eleganckim raporcie rocznym. W tym kontekście własność danych urasta do rangi kluczowego problemu ustrojowego, dotykającego samych fundamentów naszej wolności w XXI wieku. Czy nasze najbardziej intymne informacje mają być jedynie surowcem wydobywanym przez prywatne podmioty, czy może elementem infrastruktury dobra publicznego? W medycynie ten dylemat jest szczególnie jaskrawy, gdyż dane zdrowotne, choć ratują życie, mogą stać się narzędziem ubezpieczeniowej dyskryminacji lub biologicznej segregacji. Aidealizm postuluje zatem, by pacjent przestał być jedynie pasywnym dostawcą surowca dla farmaceutyczno-algorytmicznego kompleksu, stając się realnym współwłaścicielem wytwarzanej wartości.

Wymaga to jednak przejścia od czysto formalnej, bezrefleksyjnej zgody do rzeczywistej kontroli nad własnym cyfrowym śladem, co jest warunkiem zachowania godności. Dzisiejsza zgoda użytkownika jest bowiem w dużej mierze fikcją, a kliknięcie w regulamin rzadko bywa aktem autonomii, skoro alternatywą jest cyfrowe wykluczenie. Język tych dokumentów przypomina raczej hermetyczny rytuał inicjacyjny dla prawników cierpiących na bezsenność niż jasną umowę społeczną między partnerami. Musimy zatem wspierać modele powiernicze i spółdzielcze, w których dane są zarządzane przez instytucje zobowiązane do działania na rzecz osób, których te informacje dotyczą. Dane wymagają nie tylko ochrony prawnej, lecz przede wszystkim nowej konstytucji własnościowej, która zdefiniuje na nowo relacje między jednostką a systemem.

Tę kwestię można ująć krótko: kto posiada dane, ten nie tylko dysponuje wiedzą o naszej przeszłości, lecz przede wszystkim negocjuje naszą wspólną przyszłość. Dane są bowiem paliwem predykcji, a predykcja to nic innego jak władza nad prawdopodobieństwem podejmowanych przez nas decyzji. Jeśli pozwolimy, by ta władza została całkowicie sprywatyzowana, zrezygnujemy z resztek naszej podmiotowości na rzecz algorytmicznego zarządcy. Cywilizacja nie ginie najczęściej dlatego, że zabrakło jej narzędzi; ginie dlatego, że narzędzia przeżyły jej sumienie, pozostawiając nas w świecie doskonałych wyników i pustych znaczeń. Właśnie dlatego walka o własność danych jest w istocie walką o to, czy w świecie zdominowanym przez maszyny pozostanie jeszcze miejsce na ludzką wolność.

AI jako Anty-Machiavel: między nadzorem a technokracją

Jeśli korporacja wie, czego będziesz pragnąć, zanim sam nadasz temu pragnieniu nazwę, nie sprzedaje ci już produktu w tradycyjnym sensie; sprzedaje ci wersję samego siebie, którą wcześniej skrupulatnie obliczyła. Ta subtelna ewolucja od profilowania konsumenta do modelowania obywatela sprawia, że ekonomia AI płynnie i niemal niezauważalnie przechodzi w politykę AI. Skoro sztuczna inteligencja koncentruje bogactwo z niespotykaną dotąd precyzją, to z tą samą siłą musi koncentrować wpływ, czyniąc z modelowania opinii publicznej najpotężniejsze narzędzie współczesnej władzy. Aidealizm, czyli filozoficzna rama będąca odpowiedzią na współczesne kryzysy prawdy, pracy i demokracji, wybiera jednak inny kierunek: AI jako anty-Machiavela. Figura ta jest o tyle fascynująca, że obnaża główne napięcie polityki, w której technologia może stać się albo klatką, albo lustrem prawdy.

Machiavelli symbolizuje władzę rozumianą jako czystą sztukę skuteczności, gdzie moralność jest jedynie zmienną podporządkowaną celowi utrzymania panowania nad masami. Anty-Machiavel odwraca tę logikę, postulując, że władza nie ma być sztuką oszustwa, lecz przejrzystą służbą wobec dobra wspólnego. W epoce krzemu oznacza to systemy zdolne do wykrywania korupcji, mapowania ukrytych sieci wpływów i śledzenia przepływów finansowych z dokładnością, której nie powstydziłby się najlepszy śledczy. Wyobraźmy sobie algorytmy oznaczające deepfake'i, chroniące sygnalistów i symulujące skutki ustaw w sposób jawny dla każdego obywatela jeszcze przed ich uchwaleniem. Brzmi to pociągająco, wręcz utopijnie, ale właśnie tutaj musimy zachować czujność graniczącą z paranoją, bo technologia nie posiada własnego sumienia.

Anty-Machiavel może bowiem bardzo łatwo zostać Machiavellim z aktualizacją oprogramowania i nieco bardziej eleganckim interfejsem użytkownika. System wykrywający korupcję może zostać

użyty selektywnie przeciwko opozycji, a narzędzie do walki z dezinformacją może w mgnieniu oka stać się cyfrowym cenzorem. Istnieje realne ryzyko, że zamiast sprawiedliwości otrzymamy automatycznego Robespierre'a, który w imię cnoty klasyfikuje obywateli według nieprzejrzystego wskaźnika podejrzaności. Dlatego polityczna AI musi być oparta na zasadzie niezależnego nadzoru, pluralizmu i bezwzględnej jawności procedur, do których każdy ma prawo wglądu. Nie może być ona własnością rządu bez kontroli, korporacji bez odpowiedzialności ani tajnej agencji, która nie uznaje żadnych granic prawnych.

Aidealizm trafnie odrzuca niebezpieczną ideę zastąpienia polityków przez algorytmy, ponieważ demokratyczna decyzja nigdy nie jest wyłącznie operacją optymalizacyjną. Jest ona aktem odpowiedzialności, konfliktu wartości i wyrazem legitymacji, której maszyna, choćby najdoskonalsza, nigdy nie zdoła samodzielnie wygenerować. Model może wskazać, że określona polityka mieszkaniowa zmniejszy nierówności, ale nie może sam rozstrzygnąć, jaki poziom redystrybucji jest moralnie akceptowalny dla społeczeństwa. Może symulować skutki reformy podatkowej, ale nigdy nie poniesie odpowiedzialności przed wyborcami za ich puste portfele i zawiedzione nadzieje. Algorytm nie poda się do dymisji; co najwyżej otrzyma patch, który po cichu naprawi błąd, nie naprawiając jednak wyrządzonej ludziom krzywdy.

Demokracja wspierana przez AI jest więc znacznie lepszą formułą niż demokracja przez AI zastąpiona, gdyż maszyna ma analizować, a człowiek decydować. Zasada human-in-the-loop, czyli wymóg obecności człowieka w pętli decyzyjnej, nie jest sentymentalnym przywiązaniem do naszych biologicznych niedoskonałości. Jest ona fundamentalnym warunkiem politycznej odpowiedzialności, bez której nie istnieje wina, skrucha ani możliwość realnego rozliczenia władzy przez suwerena. Bez człowieka w pętli system staje się bezdusznym mechanizmem, w którym błąd jest jedynie statystyką, a nie osobistą tragedią konkretnego obywatela. Technologiczny libertarianizm zbyt często przypomina człowieka, który wjechał windą na szczyt wieżowca, a potem ogłosił, że schody są dla słabych.

Jednym z najbardziej kontrowersyjnych elementów tej wizji jest koncepcja Psychopath Power Takeover Prophylaxis, czyli profilaktyki przejęcia władzy przez osoby o cechach psychopatycznych. Intuicja stojąca za tym pomysłem jest bolesna, ale zrozumiała, jeśli spojrzymy na krwawe lekcje historii XX wieku. Widzieliśmy aż nadto osobowości narcystycznych, paranoicznych i sadystycznych, które doprowadziły całe narody do masowych katastrof i moralnego upadku. Skoro społeczeństwo wymaga konkretnych kwalifikacji od chirurga czy pilota, to dlaczego nie miałoby wymagać minimalnej zdolności moralnej od ludzi decydujących o wojnie i pokoju? A jednak to właśnie tutaj Aidealizm musi zachować największą dyscyplinę prawną, by nie stać się narzędziem nowej, naukowej tyranii.

Profilowanie psychologiczne kandydatów do władzy może stać się narzędziem ochrony demokracji, ale równie dobrze może posłużyć do jej ostatecznego zniszczenia. Każdy reżim autorytarny z radością ogłosi swoich przeciwników osobowościami niebezpiecznymi, zamieniając psychologię w pałkę retoryczną do uciszania wszelkiej krytyki. Dlatego ewentualna profilaktyka nie może polegać na tajnych, algorytmicznych diagnozach eliminujących kandydatów z wyścigu o głosy obywateli. Potrzebujemy raczej instytucjonalnej immunologii demokracji, czyli wzmocnienia jawności zachowań publicznych, audytów decyzji i edukacji obywatelskiej dotyczącej cech autorytarnego przywództwa. AI może analizować wzorce komunikacji i dehumanizację przeciwników, ale nie powinna samodzielnie wydawać quasi-klinicznych wyroków o ludzkiej osobowości.

Aidealizm jest przekonujący również dlatego, że nie sprowadza ryzyka związanego z AI do jednego, filmowego scenariusza zagłady ludzkości. Zagrożenia są wielopoziomowe i zaczynają się od manipulacji samą tkanką rzeczywistości, co podważa podstawy zaufania społecznego. Deepfake'i i syntetyczne kampanie dezinformacyjne sprawiają, że instytucje dowodowe demokracji i sądownictwa zostają drastycznie osłabione w swoich fundamentach. Powraca tu przestroga Hannah Arendt: celem permanentnego kłamstwa nie jest to, by ludzie uwierzyli w konkretną fikcję, lecz by przestali wierzyć w możliwość odróżnienia prawdy od fałszu. W świecie, w którym wszystko może być sfabrykowane, obywatel staje się bezbronny wobec siły, która kontroluje narrację i technologię.

Drugie ryzyko to automatyzacja decyzji krytycznych w obszarach takich jak medycyna, wojsko czy sądownictwo, gdzie błąd maszyny ma skutki nieodwracalne. Nie dzieje się tak dlatego, że maszyna jest kapryśna, lecz dlatego, że może być nadmiernie skuteczna w realizacji źle określonego celu. System optymalizujący koszty leczenia może po cichu ograniczać dostęp do terapii osobom starszym, a system optymalizujący sukces militarny może niebezpiecznie obniżyć próg użycia przemocy. Potwór cyfrowy nie musi ryczeć; może mieć dashboard, model ryzyka i elegancki raport zgodności, który maskuje moralną pustkę podjętej decyzji. To właśnie tutaj brak sumienia maszyny staje się jej najbardziej przerażającą i niszczycielską cechą.

Trzecie ryzyko to nadzór, czyli budowa cyfrowego panoptykonu, który potrafi łączyć dane o naszych twarzach, lokalizacji i relacjach w jedną, totalną bazę. Cyfrowy strażnik jest groźniejszy od klasycznego, ponieważ jest rozproszony w kamerach, modelach predykcyjnych i procedurach dostępu, nie potrzebując już wieży strażniczej. W państwie prawa te możliwości wymagają ścisłej kontroli sądowej, ale w rękach autokraty stają się spełnionym marzeniem każdego aparatu bezpieczeństwa. Nieprzejrzystość egzystencjalna, czyli prawo do posiadania sfery życia wolnej od algorytmicznej interpretacji, staje się w tym kontekście kluczowym elementem ochrony ludzkiej godności przed totalną kontrolą rynku lub państwa.

Czwartym zagrożeniem są uprzedzenia algorytmiczne, będące w istocie cyfrowym dziedzictwem naszych historycznych nierówności i instytucjonalnych skrzywień. Modele uczą się na danych, które są zapisem dyskryminacji, więc jeśli system rekrutacyjny trenuje się na historii branży zdominowanej przez jedną grupę, będzie on tę dominację reprodukował. Algorytm nie musi nienawidzić, żeby dyskryminować; wystarczy, że statystycznie odziedziczy świat taki, jaki był, i uzna go za jedyny możliwy wzorzec. W ten sposób uprzedzenia z przeszłości zostają zabetonowane w przyszłości pod pozorem neutralnej, matematycznej predykcji, co czyni walkę o sprawiedliwość społeczną jeszcze trudniejszą i bardziej sformalizowaną.

Piąte ryzyko to problem „czarnej skrzynki", czyli brak wyjaśnialności decyzji podejmowanych przez coraz bardziej złożone i nieprzejrzyste modele sztucznej inteligencji. W obszarach praw podstawowych brak jasnego uzasadnienia jest naruszeniem godności proceduralnej, bo człowiek ma prawo wiedzieć, dlaczego odmówiono mu kredytu czy wolności. Decyzja niezrozumiała jest decyzją quasi-feudalną, w której pan systemu posiada wiedzę, a poddany jedynie doświadcza skutków jego nieodgadnionej woli. Szóste ryzyko to gigantyczna koncentracja władzy w rękach kilku cyfrowych gigantów, którzy kontrolują architekturę naszej codzienności. Gdy siedem z dziesięciu najcenniejszych firm świata to korporacje technologiczne, nie mówimy już o sektorze gospodarki, lecz o nowej, globalnej strukturze suwerenności.

Ostatnie ryzyko dotyczy środowiska i infrastruktury, o których często zapominamy, patrząc na eleganckie i sterylne interfejsy naszych urządzeń. Ogromne zużycie energii i zasobów potrzebnych do zasilania modeli AI stanowi realne wyzwanie dla naszej ekologicznej stabilności i przetrwania. Aidealizm przypomina nam, że technologia, która niszczy własne siedlisko, nie jest postępem, lecz formą zbiorowego zaślepienia i cywilizacyjnego regresu. Potrzebujemy koncepcji takiej jak powszechna dywidenda AI, by sprawiedliwie dzielić korzyści z automatyzacji, nie niszcząc przy tym fundamentów planety. Cywilizacja nie ginie najczęściej dlatego, że zabrakło jej narzędzi; ginie dlatego, że narzędzia przeżyły jej sumienie, a my nie potrafiliśmy ich w porę okiełznać. Barbarzyństwo z interfejsem premium nadal pozostaje barbarzyństwem, nawet jeśli ubrane jest w szaty najnowocześniejszej technologii.

Wielowarstwowa obrona: jak ustrzec AI przed nadużyciami

Centra danych, te rzekomo eteryczne „chmury", pożerają energię z żarłocznością, która zawstydza tradycyjne gałęzie przemysłu. Skoro zapotrzebowanie na prąd generowane przez systemy AI rośnie szybciej niż w jakimkolwiek innym sektorze gospodarki, musimy przestać postrzegać rozwój

technologii jako izolowaną wyspę innowacji. W dobie idealizmu polityka cyfrowa staje się nierozerwalnie spleciona z polityką energetyczną, przemysłową oraz klimatyczną, tworząc jeden gęsty węzeł interesów. Nie da się budować inteligentnej przyszłości na fundamencie ekologicznego długu, którego nie zamierzamy spłacić. Potwór cyfrowy nie musi ryczeć; wystarczy, że cicho drenuje nasze wspólne zasoby, ukryty za eleganckim dashboardem i raportem zgodności.

Najwięcej emocji budzi jednak ósme ryzyko, czyli widmo autonomicznej ogólnej inteligencji, która mogłaby wymknąć się spod ludzkiej kontroli. W środowisku badaczy brakuje konsensusu: jedni wieszczą rychły koniec antropocenu, inni zaś drwią, że dzisiejsze modele są równie bliskie świadomości, co biurowy kalkulator. Idealizm proponuje tu postawę roztropnej asymetrii, zakładając, że nawet przy nikłym prawdopodobieństwie katastrofy jej potencjalny koszt jest zbyt wysoki, by go ignorować. Niemniej jednak lęk przed mitycznym Skynetem nie może stać się zasłoną dymną dla realnych, dzisiejszych patologii. Monopole, dyskryminacja algorytmiczna czy wszechobecna inwigilacja to pożary, które płoną tu i teraz, podczas gdy my debatujemy o teoretycznym wybuchu wulkanu.

Odpowiedzią na tę złożoność jest koncepcja wielowarstwowej obrony, która zaczyna się od fundamentu technicznego. Nie wystarczy deklarować bezpieczeństwa; trzeba je wymusić poprzez wyjaśnialność, rygorystyczne testy odporności oraz audyty przeprowadzane przez niezależne podmioty. Bezpieczne projektowanie oznacza tu nakładanie twardych ograniczeń na autonomię systemów i wdrażanie mechanizmów awaryjnych, które zadziałają nawet wtedy, gdy logika modelu zawiedzie. Monitorowanie systemów po ich wdrożeniu staje się koniecznością, a nie opcją, bo algorytm nie poda się do dymisji, co najwyżej otrzyma patch po fakcie. Technologia bez wbudowanych hamulców bezpieczeństwa jest jedynie zaproszeniem do chaosu, ubranym w szaty postępu.

Druga warstwa, prawna, musi nadać tym technicznym bezpiecznikom moc powszechnie obowiązującego imperatywu. Obejmuje ona licencjonowanie systemów wysokiego ryzyka, określenie odpowiedzialności producentów oraz bezwzględny zakaz praktyk niedopuszczalnych, godzących w godność jednostki. Prawo do wyjaśnienia decyzji podjętej przez maszynę staje się nowym prawem człowieka, chroniącym nas przed arbitralnością kodu. Konieczne są również międzynarodowe traktaty, które wyjmą systemy AI spod logiki wyścigu zbrojeń i dezinformacyjnych wojen hybrydowych. Bez twardych ram prawnych etyka w technologii pozostaje jedynie dobrowolną obietnicą, którą można wycofać przy pierwszym spadku kwartalnych zysków.

Trzecim filarem jest warstwa społeczna, bez której nawet najdoskonalsze paragrafy pozostaną martwą literą na papierze. Kluczowa staje się edukacja w zakresie AI oraz budowanie odporności informacyjnej, która pozwoli obywatelom odróżnić syntetyczną prawdę od rzeczywistości. Musimy

tworzyć publiczne alternatywy dla komercyjnych modeli, aby uniknąć cyfrowego feudalizmu, w którym infrastruktura myślowa społeczeństwa należy do kilku korporacji. Kontrola obywatelska i kultura krytycznego myślenia to jedyne skuteczne antidotum na niedojrzałość cyfrową, czyli stan bierności, w którym rezygnujemy z własnego sądu na rzecz algorytmu. Tylko świadome społeczeństwo potrafi patrzeć maszynie na ręce, zamiast bezkrytycznie przyjmować jej werdykty.

Prawdziwa siła tej obrony tkwi jednak nie w poszczególnych warstwach, lecz w ich nierozerwalnej, organicznej synergii. Techniczne zabezpieczenie bez wsparcia prawa jest jedynie uprzejmą prośbą, a prawo bez kompetencji społecznej staje się biurokratyczną fasadą. Edukacja bez kontroli nad ekonomią platform przypomina naukę pływania w rzece, której nurt zaprojektowano tak, by nieuchronnie niósł wszystkich w stronę kasyna. Wielowarstwowa obrona musi działać jako system wzajemnych hamulców, ponieważ AI jest technologią zbyt potężną, by ufać jednej barierze i modlić się do dokumentacji. Barbarzyństwo z interfejsem premium nadal pozostaje barbarzyństwem, jeśli jedynym zabezpieczeniem jest nasza wiara w dobrą wolę programistów.

Rzetelna obrona idealizmu wymaga jednak odwagi w punktowaniu jego własnych słabych punktów, z których pierwszym jest ryzyko technokratycznego idealizmu. Jeśli zaczniemy postrzegać AI jako neutralnego arbitra racjonalności, łatwo przeoczymy fakt, że każdy model jest nasycony wartościami swoich twórców i instytucji. „Obiektywna AI" to niebezpieczny mit, który służy legitymizacji władzy tych, którzy trzymają rękę na pulsie serwerowni. Zamiast gonić za niemożliwą neutralnością, powinniśmy dążyć do AI audytowalnej, pluralistycznej i świadomej własnych ograniczeń. Algorytm nie jest głosem Boga, lecz skondensowanym echem naszych własnych, często wadliwych założeń i uprzedzeń.

Druga wątpliwość dotyczy samego pojęcia dobra wspólnego, którym idealizm tak chętnie operuje w swoich manifestach. Musimy przyznać, że pojęcia takie jak sprawiedliwość czy demokracja są areną nieustannego sporu, a nie gotowym zestawem danych do zoptymalizowania. Inaczej zdefiniuje je liberalny konstytucjonalista, inaczej socjaldemokrata, a jeszcze inaczej zwolennik wspólnotowości czy ruchy postkolonialne. AI nie rozwiąże naszych konfliktów wartości; może je co najwyżej lepiej naświetlić, ale nigdy nie zastąpi politycznego procesu ich uzgadniania. Dlatego idealizm musi pozostać proceduralny i otwarty, broniąc warunków uczciwego sporu, zamiast narzucać jeden algorytmiczny katechizm dobra.

Trzecia słabość to widmo globalnej nierówności i obawa, że zachodnia wizja etycznej AI zostanie odebrana jako kolejna forma normatywnego eksportu. Jeśli bogate państwa Północy, które już zmonopolizowały kapitał i dane, będą dyktować reguły gry reszcie świata, idealizm stanie się luksusowym humanizmem dla wybranych. Prawdziwie inkluzywna technologia wymaga transferu wiedzy, dostępu do infrastruktury publicznej i przeciwdziałania cyfrowemu kolonializmowi.

Musimy dbać o to, by kraje rozwijające się nie były jedynie dostarczycielami surowych danych i tanią siłą roboczą do trenowania modeli. Bez sprawiedliwego podziału korzyści AI pogłębi jedynie przepaść między tymi, którzy projektują przyszłość, a tymi, którzy są na nią skazani.

Wreszcie nie wolno lekceważyć głosów krytyków, którzy w regulacjach widzą jedynie hamulec dla innowacji i osłabienie konkurencyjności wobec autorytarnych potęg. To realna obawa, niemniej odpowiedź idealizmu musi być w tej kwestii nieugięta: nie każda przewaga technologiczna zasługuje na miano cywilizacyjnego zwycięstwa. Jeśli demokracje, chcąc wygrać wyścig z autokracjami, same zbudują infrastrukturę totalnego nadzoru i manipulacji, to wygrają technicznie, przegrywając jednocześnie swój ustrojowy sens. Cywilizacja nie ginie najczęściej dlatego, że zabrakło jej narzędzi; ginie dlatego, że narzędzia przeżyły jej sumienie. Prawdziwa siła tkwi w zdolności do samoograniczenia w imię wartości, które czynią nas ludźmi.

Aidealizm: między pokojem republikańskim a cyfrowym panoptykonem

Nie zamierzamy hamować innowacji, lecz musimy wreszcie odczarować to słowo, by przestało służyć jako uniwersalne rozgrzeszenie dla każdej formy cyfrowego drapieżnictwa. Innowacja nie może być wytrychem otwierającym drzwi do bezkarnej eksploatacji danych czy automatyzacji przemocy, która ukrywa się za eleganckim interfejsem i obietnicą optymalizacji. Piąta wątpliwość, z którą musimy się zmierzyć, dotyczy czysto technicznej wykonalności Powszechnej Dywidendy AI, czyli koncepcji sprawiedliwego podziału korzyści ekonomicznych płynących z automatyzacji i wzrostu produktywności maszynowej, która pozwala zapobiegać drastycznym nierównościom społecznym. Pytania o to, jak precyzyjnie mierzyć zyski generowane przez modele, jak skutecznie unikać arbitrażu podatkowego czy jak odróżnić realną produktywność algorytmiczną od zwykłej cyfryzacji procesów, są jak najbardziej zasadne. Barbarzyństwo z interfejsem premium nadal pozostaje barbarzyństwem.

Trudność wdrożenia nowej architektury ekonomicznej nigdy nie stanowiła ostatecznego dowodu na jej błędność, o czym chętnie zapominają dzisiejsi sceptycy. Państwo opiekuńcze, progresywne podatki dochodowe czy systemy ubezpieczeń społecznych również wydawały się utopijnymi konstruktami, zanim stały się fundamentami nowoczesnych społeczeństw. Argument głoszący, że „to zbyt skomplikowane”, jest zazwyczaj jedynie elegancką wersją wezwania do ochrony interesów najsilniejszych graczy rynkowych. Ekonomia sztucznej inteligencji staje się dziś polem bitwy, na którym aidealizm, czyli filozoficzna rama będąca odpowiedzią na kryzysy prawdy, pracy, demokracji, ekologii oraz antropologii, albo przekuje się w realną filozofię ustrojową, albo

pozostanie jedynie efektownym słownikiem. Algorytm nie poda się do dymisji, co najwyżej otrzyma patch, dlatego to my musimy zaprojektować ramy, w których będzie on funkcjonował.

Jeśli sztuczna inteligencja pozostanie własnością nielicznych, a dane surowcem wydobywanym bez udziału obywateli, czeka nas powrót do struktur feudalnych. Jeśli automatyzacja ostatecznie zerwie więź między produktywnością a społecznym dobrobytem, obietnica Oświecenia zostanie zdradzona w sposób wyjątkowo ironiczny: przez maszyny będące owocem czystego rozumu. Aidealizm proponuje jednak inną drogę, w której AI staje się infrastrukturą dobra wspólnego, a nie narzędziem bezwzględnej akumulacji kapitału. Dane wymagają tu statusu zasobu objętego powiernictwem i współwłasnością, co chroniłoby je przed totalną komercjalizacją i nadzorem. Powszechna Dywidenda AI jawi się w tym kontekście jako jedyna racjonalna odpowiedź na automatyzację wartości, która inaczej bezpowrotnie wyparuje z portfeli obywateli.

W tej wizji anty-Machiavelli nie jest cenzorem, lecz narzędziem radykalnej przejrzystości, pozwalającym patrzeć władzy na ręce, nawet jeśli ta władza nie ma ludzkiej twarzy. Potrzebujemy wielowarstwowej obrony, systemu hamulców i równowagi, który powstrzyma technologię przed staniem się wszechobecnym, cyfrowym lewiatanem. Najkrótsza formuła brzmi: sztuczna inteligencja będzie sprawiedliwa tylko wtedy, gdy sprawiedliwe będą instytucje kontrolujące jej własność i dystrybucję. Sama moc obliczeniowa nie posiada wrodzonego genu solidarności, jest ona etycznie obojętna i podatna na każdą formę nadużycia. Solidarność trzeba jej zatem narzucić siłą prawa, głębią kultury oraz nieustanną, obywatelską czujnością, która nie daje się uśpić obietnicą wygody.

Podczas gdy ekonomia AI pyta o to, kto przejmie nadwyżkę z automatyzacji, polityczna AI drąży kwestię kontroli nad władzą ukrytą w czarnych skrzynkach modeli. Kantowska perspektywa wiecznego pokoju stawia te pytania jeszcze wyżej, zmuszając nas do refleksji nad przyszłością zorganizowanej przemocy. Czy sztuczna inteligencja pomoże ludzkości wyjść z epoki wojen, czy przeciwnie – uczyni konflikt szybszym, bardziej autonomicznym i poznawczo nieuchwytnym dla obywateli? Kant nie był naiwnym pacyfistą w cukrowej peruce metafizyki, lecz architektem porządku prawnego, w którym wojna staje się anomalią wobec rozumu. Jego projekt nie polegał na sentymentalnych życzeniach, lecz na stworzeniu struktury wzajemnych ograniczeń, która czyni agresję nieopłacalną.

Pokój nie jest jedynie chwilowym zawieszeniem broni, lecz trwałą formą instytucjonalnej dojrzałości, która wyklucza przemoc jako narzędzie prowadzenia polityki. W epoce wszechobecnych algorytmów świat może zostać „uspokojony” na dwa zasadniczo odmienne sposoby, które zdefiniują naszą przyszłość. Pierwszy to pokój republikański, oparty na jawności, prawie i wzajemnym uznaniu godności, co wymaga od nas ogromnego wysiłku organizacyjnego. Drugi to

pokój panoptyczny, bazujący na totalnej predykcji, neutralizacji oporu i prewencyjnym zarządzaniu ryzykiem społecznym przez systemy nadzoru. Pierwszy z nich tworzy świadomych obywateli, podczas gdy drugi produkuje jedynie czytelne dla systemu obiekty administracji, pozbawione realnej podmiotowości.

Aidealizm musi jednoznacznie opowiedzieć się za modelem republikańskim, odrzucając wizję AI jako cyfrowego żandarma pilnującego globalnego status quo. Jeśli technologia ta ma być strażnikiem pokoju, nie może stać się narzędziem, które w imię bezpieczeństwa likwiduje resztki ludzkiej wolności. Pokój pozbawiony godności nie jest w rzeczywistości pokojem, lecz jedynie stabilizacją uległości w systemie, który nie dopuszcza błędu ani sprzeciwu. Cmentarz również bywa miejscem spokojnym, jednak nikt nie uznałby go za wzór idealnego ustroju politycznego. Musimy zatem wybrać między żywym dialogiem wolnych ludzi a martwą ciszą doskonale zoptymalizowanego algorytmu, który nie zna litości, bo nie zna sumienia.

Algorytmiczny pokój: między wojną poznawczą a nową prawdą

Kantowska wizja wiecznego pokoju, sformułowana u progu nowoczesności, opierała się na eleganckim, niemal mechanicznym założeniu ludzkiej racjonalności. Filozof z Królewca wierzył, że trwała stabilizacja między narodami wymaga republikańskich konstytucji, w których to obywatele, a nie monarchowie, decydują o wszczęciu konfliktu. Logika była bezlitosna w swej prostocie: jeśli to ty musisz płacić podatki krwią, zdrowiem i majątkiem, dwa razy zastanowisz się, zanim chwycisz za broń. W epoce modeli predykcyjnych ta intuicja wymaga jednak bolesnej aktualizacji, ponieważ technologia pozwala na radykalne oddzielenie realnych kosztów wojny od jej społecznej widzialności. Barbarzyństwo z interfejsem premium nadal pozostaje barbarzyństwem, nawet jeśli krew nie plami ekranów naszych smartfonów w czasie rzeczywistym.

Dzisiejsze pole bitwy nie przypomina już frontów znanych z map sztabowych, lecz raczej rozproszoną sieć impulsów, w której granica między pokojem a agresją ulega całkowitemu zatarciu. Autonomiczne systemy wojskowe, precyzyjne cyberoperacje oraz syntetyczna propaganda pozwalają prowadzić działania destrukcyjne znacznie poniżej progu klasycznej deklaracji wojny. Wojna staje się mglista, hybrydowa i częściowo zaprzeczalna, co sprawia, że tradycyjne mechanizmy demokratycznej kontroli stają się niepokojąco bezradne. Państwo może dziś skutecznie paraliżować infrastrukturę przeciwnika lub manipulować jego procesami wyborczymi, nie wysyłając ani jednego żołnierza przez fizyczną granicę. W takich warunkach kantowskie pytanie

o zgodę obywateli traci swój fundament, bowiem obywatel często nie ma pojęcia, że od dawna znajduje się w samym centrum strefy rażenia.

Sztuczna inteligencja występuje w tym dramacie w roli wybitnie dwuznacznej, będąc jednocześnie podpałką i gaśnicą dla globalnych napięć. Z jednej strony może służyć jako potężne narzędzie agresji, automatyzując generowanie dezinformacji i mikrotargetowanie lęków społecznych w celu destabilizacji kruchych instytucji. Z drugiej strony AI może stać się instrumentem pokoju, jeśli wykorzystamy ją do wczesnego wykrywania eskalacji oraz monitorowania naruszeń prawa międzynarodowego w sposób obiektywny. Tutaj właśnie pojawia się aidealizm, czyli filozoficzna rama i odpowiedź na współczesne kryzysy prawdy, pracy oraz demokracji, która nakazuje nam widzieć w technologii szansę na odzyskanie przejrzystości świata. Algorytm nie poda się do dymisji, ale może zmusić polityków do konfrontacji z faktami, których nie da się już tak łatwo zignorować w blasku algorytmicznej prawdy.

Modele predykcyjne, choć potrafią z imponującą dokładnością ostrzegać przed nadchodzącym kryzysem, nigdy nie zastąpią ludzkiej decyzji o charakterze czysto moralnym. Pokój nie jest bowiem prostym wynikiem matematycznego równania, lecz świadomym wyborem politycznym, wymagającym odwagi do negocjacji, przebaczenia lub ryzykownej obrony wspólnych wartości. Niemniej jednak AI może znacząco zmniejszyć obszar systemowej ignorancji, w której zwykle dojrzewają najbardziej krwawe i bezsensowne konflikty naszej historii. Może ona demaskować fałszywe preteksty do inwazji oraz brutalnie obnażać koszty, których wojenna propaganda nigdy nie odważyłaby się nazwać po imieniu przed własnym społeczeństwem. Potwór cyfrowy nie musi ryczeć; może po prostu dostarczyć nam raport, który odbierze wojnie jej ulubionego sojusznika: dobrze zorganizowaną niewiedzę.

Przejście od wojny żelaza do wojny poznania oznacza, że dzisiejsze konflikty toczą się przede wszystkim o kontrolę nad ludzką percepcją i zdolnością do odróżniania dowodu od symulacji. Dawna propaganda, oparta na topornych rytuałach masowego przekazu, została zastąpiona przez modele generatywne zdolne do tworzenia nieskończonej liczby syntetycznych tożsamości i narracji. Wojna poznawcza nie dąży już do przekonania nas do konkretnej ideologii, lecz do rozbicia społeczeństwa na plemiona pogrążone we wzajemnej, paraliżującej nieufności. Kiedy znikają wspólne punkty odniesienia, przestrzeń publiczna zamienia się w toksyczne bagnisko, w którym każda informacja jest traktowana jako potencjalna pułapka zastawiona przez wroga. W takim świecie prawda przestaje być celem samym w sobie, a staje się jedynie kolejnym zasobem do zmanipulowania w imię doraźnego zysku politycznego.

W tym kontekście diagnozy Hannah Arendt nabierają przerażającej aktualności, ostrzegając nas przed społeczeństwem, które przestało wierzyć w możliwość obiektywnego poznania rzeczywistości.

Niedojrzałość cyfrowa, czyli stan bierności intelektualnej, w którym rezygnujemy z krytycznego myślenia na rzecz gotowych odpowiedzi algorytmicznych, czyni nas idealnymi poddanymi dla nowoczesnych systemów manipulacji. Cynik polityczny uważa się za odpornego na propagandę, podczas gdy w rzeczywistości ufa każdemu komunikatowi, który skutecznie żeruje na jego gniewie i potwierdza jego najgorsze uprzedzenia. Aidealizm musi zatem traktować integralność informacyjną jako absolutny warunek wstępny jakiegokolwiek trwałego pokoju między ludźmi i narodami. Bez minimalnego zaufania do dowodów i faktów demokracja zamienia się w pustą procedurę, a prawo w teatr cieni rzucanych przez cudze, nieprzejrzyste modele.

Ochrona warunków prawdy w epoce deepfake'ów wymaga od nas budowy nowej, technologicznie świadomej infrastruktury wiarygodności, która wykracza poza reaktywne gaszenie pożarów dezinformacji. Nie chodzi tu o wprowadzenie cenzury, lecz o stworzenie systemów autentykacji treści, które pozwolą nam odróżnić głos człowieka od automatycznie wygenerowanego szumu maszynowego. Oznaczanie materiałów syntetycznych oraz penalizacja złośliwych manipulacji w procesach wyborczych to nie zamach na wolność słowa, lecz próba ocalenia samej możliwości sensownej wypowiedzi w sferze publicznej. Wolność nie polega bowiem na prawie do zalania świata cyfrowym śmieciem, który niszczy tkankę społeczną i uniemożliwia jakąkolwiek racjonalną debatę o przyszłości. Jeśli nie zabezpieczymy fundamentów prawdy, sprawiedliwość będzie sądzić jedynie cyfrowe fantomy, a nie realne czyny i ich rzeczywistych sprawców.

Spojrzenie na te problemy z perspektywy antropologii ewolucyjnej pozwala zrozumieć, dlaczego jako gatunek czujemy się tak zagubieni w świecie, który sami sobie zbudowaliśmy. Nasz układ nerwowy, ukształtowany tysiące lat temu w zupełnie innych warunkach, reaguje błyskawicznie na bezpośrednie zagrożenia, ale pozostaje niemal ślepy na abstrakcyjne ryzyka systemowe. Dezinformacja nie ma twarzy drapieżnika, a kryzysy instytucjonalne nie nadbiegają z włócznią, przez co nasze reakcje obronne są często spóźnione lub skierowane przeciwko niewłaściwym bodźcom. AI może jednak stać się rozszerzeniem naszej ograniczonej racjonalności, pomagając nam dostrzec wzorce i zagrożenia, które umykają naszej ułomnej, biologicznej intuicji. Technologiczny libertarianizm zbyt często przypomina człowieka, który wjechał windą na szczyt wieżowca, a potem ogłosił wszem i wobec, że schody są tylko dla słabych.

Połączenie sztucznej inteligencji z nowoczesną genomiką i medycyną precyzyjną otwiera przed nami perspektywę wyjścia poza biologiczne ograniczenia, które dotąd definiowały ludzki los. Możemy dziś modelować długoterminowe skutki naszych działań, diagnozować choroby na lata przed ich wystąpieniem i projektować interwencje zdrowotne o niespotykanej dotąd skuteczności. Nierówność antropotechniczna, czyli potencjalny podział ludzkości wynikający z nierównego dostępu do ulepszeń biologicznych i poznawczych, pozostaje jednak realnym zagrożeniem, któremu

musimy aktywnie przeciwdziałać. AI nie jest magiczną różdżką, która sama zaprowadzi nas do technologicznej utopii, lecz potężnym narzędziem, które wymaga od nas nowej formy dojrzałości i sumienia. Cywilizacja nie ginie najczęściej dlatego, że zabrakło jej odpowiednich narzędzi; ginie dlatego, że narzędzia te ostatecznie przeżyły jej sumienie.

Aidealizm: Instytucje, Solidarność i Człowiek w Pętli

Obietnica, którą składa nam współczesna technologia, jest doprawdy oszałamiająca i nie wolno jej lekceważyć wyłącznie z lęku przed potencjalnymi nadużyciami. Redukcja cierpienia, owa odwieczna misja medycyny i nauki, stanowi wszak fundamentalne dobro moralne, którego nie sposób zakwestionować bez popadania w jałowy nihilizm. Niemniej jednak ta sama świetlista obietnica rzuca niezwykle długi i gęsty cień na naszą wspólną, gatunkową przyszłość. Jeśli bowiem dostęp do ulepszeń biologicznych, poznawczych i zdrowotnych zostanie ostatecznie sprywatyzowany, ludzkość może niepostrzeżenie wejść w fazę głębokiej nierówności antropotechnicznej. Nierówność antropotechniczna, czyli potencjalny podział ludzkości wynikający z nierównego dostępu do sprywatyzowanych ulepszeń biologicznych i poznawczych, wykracza daleko poza ramy tradycyjnej stratyfikacji majątkowej. Dotychczas bogaci mogli kupować lepsze domy, elitarne szkoły czy poczucie bezpieczeństwa, co mieściło się jeszcze w granicach klasycznej dynamiki społecznej. W nadchodzącej erze mogą oni jednak nabywać lepszą odporność, radykalnie dłuższe życie oraz selekcyjne ulepszenia reprodukcyjne, które pozostaną całkowicie niedostępne dla reszty populacji.

Taki scenariusz nie byłby już tylko kolejną wersją niesprawiedliwości ekonomicznej, lecz stanowiłby faktyczną prywatyzację ewolucji, co podważa fundament współczesnych praw człowieka. Uniwersalizm godności zakłada bowiem, że wszyscy należymy do jednej wspólnoty moralnej, niezależnie od naszych indywidualnych zdolności, posiadanego majątku czy statusu społecznego. Jeżeli jednak wyłoni się klasa ludzi biologicznie i poznawczo ulepszonych, niemal natychmiast pojawią się ideologie naturalizujące ich rzekomą, obiektywną przewagę nad resztą. Historia zna już takie języki wykluczenia, które zawsze zaczynają się od wskazywania na różnice, a kończą na brutalnej odmowie pełnego człowieczeństwa. Cyfrowy feudalizm mógłby wówczas przeobrazić się w feudalizm genomowy, gdzie właściciele kapitału staliby się jednocześnie panami przyspieszonej ewolucji biologicznej. Reszta ludzkości, pozbawiona dostępu do tych zasobów, zostałaby uznana za wersję biologiczną, która jest już beznadziejnie nieaktualna i zbędna w nowym świecie. Aidealizm, czyli filozoficzna rama i odpowiedź na pięć współczesnych kryzysów: prawdy, pracy, demokracji, ekologii oraz antropologii, musi zatem stanowczo odrzucić taką wizję przyszłości.

Jeśli sztuczna inteligencja i biotechnologia mają rzeczywiście służyć ludzkości, ich korzyści muszą zostać włączone w ramy sprawiedliwości zdrowotnej oraz globalnej solidarności. Medycyna precyzyjna nie może stać się jedynie luksusową medycyną arystokratyczną, dostępną wyłącznie dla wąskiej elity finansowej i politycznej. Dane genomowe, będące dziedzictwem całego gatunku, nie mogą być traktowane jak prywatna kopalnia surowców eksploatowana bez żadnych praw człowieka. Ulepszanie człowieka nie może wyprzedzać leczenia podstawowego, dopóki elementarna opieka medyczna pozostaje niedostępna dla milionów ludzi na całym świecie. W przeciwnym razie humanizm technologiczny stanie się jedynie salonowym, pustym słowem, służącym do eleganckiego maskowania luksusowej eugeniki rynkowej. Można to ująć brutalnie: cywilizacja, która pozwoli bogatym kupić sobie lepszą ewolucję, nie stworzy przyszłości człowieka, lecz jedynie przyszłość kastową. Barbarzyństwo z interfejsem premium nadal pozostaje barbarzyństwem, nawet jeśli jest opakowane w najnowocześniejsze rozwiązania z zakresu inżynierii genetycznej.

Aidealizm opiera się na prowokacyjnym założeniu, że ludzkość musi dopiero stać się istotą prawdziwie wartą ocalenia w obliczu własnej, technologicznej potęgi. To stwierdzenie jest celowo mocne, niebezpieczne, a zarazem rozpaczliwie potrzebne w epoce, w której nasze narzędzia wyprzedziły naszą mądrość. Jest ono mocne, ponieważ bezlitośnie odmawia nam prawa do automatycznego samozadowolenia, które tak często towarzyszy każdemu postępowi technicznemu. Bywa też niebezpieczne, gdyż może zostać instrumentalnie wykorzystane przez tych, którzy roszczą sobie prawo do oceniania, kto zasługuje na przetrwanie. Niemniej jednak pozostaje potrzebne, bo zmusza nas do postawienia pytania, czy nasz rozwój moralny nadąża za tempem innowacji. Historia nowoczesności to wszak kronika asynchronicznego postępu, w której potrafimy budować niszczycielskie narzędzia znacznie szybciej niż instytucje zdolne do ich okiełznania. Potrafimy uwalniać potężne energie, zanim nauczymy się przewidywać ich długofalowe konsekwencje, oraz produkować wiedzę, zanim wykształcimy w sobie mądrość.

Broń jądrowa była dla ludzkości pierwszym wielkim egzaminem z dojrzałości, lecz sztuczna inteligencja może okazać się wyzwaniem znacznie bardziej złożonym. Nie dotyczy ona bowiem jednego, konkretnego typu fizycznej destrukcji, lecz przenika niemal wszystkie tkanki naszych instytucji społecznych i politycznych jednocześnie. Nie chodzi tu wcale o to, że człowiek jest z natury istotą złą, gdyż takie twierdzenie byłoby zbyt proste i wygodne. Człowiek jest raczej istotą moralnie niestabilną, zdolną zarówno do wielkiego altruizmu, jak i do wyrafunowanego samozakłamania czy dominacji. Sztuczna inteligencja drastycznie zwiększy amplitudę tej niestabilności, sprawiając, że nasze małe błędy staną się wielkoskalowymi awariami systemowymi. Małe uprzedzenia, dotąd ukryte, mogą zostać zautomatyzowane i powielone w milionach wariantów, tworząc nową, cyfrową architekturę ucisku. Małe kłamstwa mogą zostać wygenerowane

w nieskończonych wersjach, a drobna chciwość może otrzymać do dyspozycji potężną, globalną infrastrukturę.

Z drugiej strony, nawet mała empatia może zostać wsparta nowoczesnymi narzędziami diagnozy, edukacji oraz sprawiedliwej redystrybucji dóbr. Mała racjonalność, którą dysponujemy, może zostać rozszerzona przez zaawansowane modele obliczeniowe, pomagając nam w rozwiązywaniu problemów o skali planetarnej. Mały akt kontroli obywatelskiej może uzyskać potężne instrumenty przejrzystości, pozwalające patrzeć na ręce tym, którzy dzierżą realną władzę. To jest właśnie sedno Aidealizmu: sztuczna inteligencja nie przesądza ostatecznego wyniku naszych zmagań, lecz jedynie drastycznie powiększa stawkę tej gry. Kantowski projekt Wiecznego Pokoju warto odczytać dziś jako postulat budowy instytucji zdolnych do skutecznego ograniczania najgorszych impulsów państw. W epoce cyfrowej instytucje te muszą obejmować nowe traktaty międzynarodowe oraz normy prawne, które nadążą za rozwojem technologii. Autonomiczna broń wymaga jasnych granic etycznych, a cyberoperacje przeciwko infrastrukturze cywilnej muszą wiązać się z realnymi sankcjami.

Syntetyczna dezinformacja wyborcza wymaga wspólnych standardów ochrony procesów demokratycznych, abyśmy nie utonęli w morzu wygenerowanego kłamstwa. Modele wysokiego ryzyka, o których mówi AI Act, czyli struktura prawna Unii Europejskiej klasyfikująca systemy według stopnia zagrożenia, wymagają współpracy. Eksport systemów nadzoru do reżimów autorytarnych powinien być traktowany nie jako zwykły handel, lecz jako współudział w architekturze represji. Nie chodzi tutaj o naiwną wiarę w globalną zgodę, gdyż państwa zawsze będą ze sobą rywalizować, a korporacje lobbować. Armie będą eksperymentować z nowymi rodzajami broni, a autokracje wykorzystywać algorytmy do totalnej kontroli nad swoimi obywatelami. Dlatego właśnie Kantowski wymiar Aidealizmu musi być głęboko realistyczny i osadzony w twardych realiach politycznych naszej współczesnej epoki. Pokój nie wynika wszak z samych dobrych intencji, lecz z mądrze zaprojektowanych ograniczeń, które krępują najgorsze ludzkie impulsy.

Moralność pozbawiona wsparcia silnych instytucji pozostaje jedynie jałowym kazaniem, podczas gdy instytucje bez fundamentu moralnego stają się bezduszną maszyną. Potrzebujemy obu tych elementów, aby przetrwać nadchodzące wstrząsy i zbudować świat, w którym technologia służy człowiekowi, a nie odwrotnie. W tym miejscu warto przywołać parafrazę Cycerona: dobro ludu jest najwyższym prawem, lecz w epoce AI musimy dodać pewne zastrzeżenie. Dobro ludu nie może być definiowane przez system, którego tenże lud nie rozumie, nie kontroluje i nie może skutecznie zaskarżyć. Salus populi nie jest bowiem zgodą na paternalizm algorytmiczny, w którym maszyna decyduje o tym, co jest dla nas najlepsze. Jest to raczej obowiązek budowy takiej infrastruktury, która wzmacnia obywatela, zamiast czynić go całkowicie przezroczystym dla władzy i kapitału.

Wieczny Pokój wymaga także zdecydowanej walki z narastającą nierównością globalną, która grozi nową formą kolonializmu technologicznego.

Państwa pozbawione własnej mocy obliczeniowej oraz kompetencji regulacyjnych staną się jedynie biernymi klientami wielkich, technologicznych imperiów rynkowych. Ich suwerenność pozostanie jedynie formalnym zapisem, podczas gdy operacyjnie będą one całkowicie uzależnione od zewnętrznych dostawców systemów i danych. Będą one zmuszone importować modele, których nie potrafią audytować, oraz wdrażać standardy, na których kształt nie miały żadnego wpływu. To byłaby nowa, niezwykle groźna forma zależności peryferyjnej, w której dane stają się surowcem wypompowywanym bez żadnej kontroli. Aidealizm jako filozofia globalna musi więc obejmować prawo do aktywnego uczestnictwa w gospodarce AI, a nie tylko bierną rolę użytkownika. Powszechna dywidenda AI, czyli koncepcja sprawiedliwego podziału korzyści ekonomicznych płynących z automatyzacji, powinna stać się fundamentem nowej umowy społecznej. Bez tego mechanizmu wzrost produktywności maszynowej doprowadzi jedynie do dalszej, niebezpiecznej koncentracji bogactwa w rękach nielicznych podmiotów.

Pojawia się często przekonanie, że Europa, jako spadkobierczyni Oświecenia, ma do odegrania szczególną rolę w budowaniu etycznego ładu cyfrowego. Ta teza brzmi niezwykle przekonująco, lecz wymaga od nas dużej dawki realizmu w ocenie własnej pozycji na mapie świata. Europa dysponuje wprawdzie silnym aparatem normatywnym i tradycją ochrony praw podstawowych, ale pozostaje w tyle w wyścigu o infrastrukturę. Może więc stać się sumieniem świata cyfrowego, lecz sumienie pozbawione realnej mocy wykonawczej bywa traktowane jedynie jako elegancki załącznik. Aidealizm europejski musi zatem umiejętnie łączyć surową regulację z odważną inwestycją w publiczną infrastrukturę sztucznej inteligencji oraz suwerenność chmurową. Nie wystarczy bowiem jedynie zakazywać pewnych praktyk, jeśli nie buduje się alternatywnych narzędzi zarządzanych w szeroko rozumianym interesie publicznym. Europa nie może być tylko notariuszem etyki, lecz musi stać się aktywnym budowniczym instytucji, które chronią godność jednostki.

Warto przy tym pamiętać, że rynek technologiczny nie jest żadnym pierwotnym stanem natury, lecz skomplikowaną i celową konstrukcją prawną. Korporacje, własność intelektualna oraz rynki kapitałowe istnieją wyłącznie dzięki ramom prawnym, które gwarantują ich bezpieczne i przewidywalne funkcjonowanie. Skoro prawo tworzy warunki do akumulacji ogromnego kapitału, ma ono również pełne prawo do określania warunków odpowiedzialności społecznej. Technologiczny libertarianizm zbyt często przypomina człowieka, który wjechał windą na szczyt wieżowca, a potem z pogardą ogłosił, że schody są dla słabych. Kto korzysta z dobrodziejstw porządku prawnego, nie może jednocześnie udawać, że każda próba regulacji jest zamachem na naturalną wolność innowacji. Bez silnego państwa prawa nie byłoby przecież bezpiecznych

kontraktów, patentów ani stabilnych rynków, na których wyrastają dzisiejsze potęgi technologiczne.

Zasada człowieka w pętli decyzyjnej powinna zostać podniesiona do rangi konstytucyjnej zasady technologicznej, stanowiącej bezpiecznik naszej cywilizacji. Nie postulujemy tego dlatego, że człowiek zawsze podejmuje decyzje lepsze, szybsze czy bardziej sprawiedliwe niż algorytmiczne modele obliczeniowe. Często bywa wręcz odwrotnie, gdyż jako istoty biologiczne decydujemy wolniej, ulegamy emocjom i wykazujemy się mniejszą konsekwencją niż maszyny. Jednakże tylko człowiek posiada unikalną zdolność do ponoszenia pełnej odpowiedzialności moralnej oraz prawnej za skutki swoich czynów. Tylko istota ludzka może zrozumieć, że sztywna reguła statystyczna w konkretnym, jednostkowym przypadku może prowadzić do rażącej niesprawiedliwości. Human-in-the-loop nie może być jednak jedynie fikcyjnym kliknięciem, które służy jako alibi dla bezdusznego systemu zarządzania algorytmicznego.

W przeciwnym razie otrzymamy zjawisko human-on-the-hook, czyli sytuację, w której człowiek jest jedynie zakładnikiem odpowiedzialności za decyzje faktycznie podjęte przez automat. Jeśli urzędnik czy lekarz nie ma czasu ani kompetencji, by zakwestionować rekomendację systemu, jego obecność staje się czystą dekoracją. Prawdziwa podmiotowość wymaga dostępu do uzasadnienia, odpowiednich zasobów czasowych oraz realnej ochrony prawnej przed presją ze strony organizacji. Jest to szczególnie krytyczne w sferze administracji publicznej, gdzie automatyczne systemy mogą decydować o przyznaniu świadczeń lub ocenie ryzyka. Bez realnej kontroli państwo prawa niebezpiecznie szybko przeobraża się w państwo scoringu, w którym obywatel staje się całkowicie przezroczysty. Aidealizm domaga się więc, aby w pętli znajdował się człowiek kompetentny, niezależny i wyposażony w skuteczne prawo do sprzeciwu. Dopiero wtedy owa zasada nabiera realnego sensu i staje się fundamentem etycznego rozwoju technologii w naszej epoce.

Aidealizm: Ostatnia próba Oświecenia w epoce algorytmów

Wyobraźmy sobie, dla lepszego zrozumienia istoty sporu, przypowieść o dwóch ogrodnikach działających w pewnym mitycznym królestwie. Pierwszy z nich, owładnięty pasją doskonałości, postanowił uczynić swój ogród absolutnie idealnym, budując w tym celu potężną maszynę do totalnego zarządzania florą. Urządzenie to z niebywałą precyzją mierzyło wilgotność każdej grudki ziemi, kąt padania promieni słonecznych oraz tempo wzrostu każdego liścia, szybko dochodząc do logicznego wniosku, że chaos jest wrogiem wydajności. Maszyna uznała więc, że najłatwiej utrzymać porządek, gdy wszystkie rośliny są identyczne, równo przycięte i w pełni przewidywalne w

swojej biologicznej egzystencji. Ogród stał się geometrycznym arcydziełem, zdrowym i sterylnym, lecz jednocześnie całkowicie martwym w swoim wewnętrznym duchu, gdyż wyeliminowano z niego wszelką niespodziankę. Pszczoły, te naturalne posłanki różnorodności, przestały tam zaglądać, nie znajdując w tej technokratycznej pustyni niczego, co warte byłoby ich trudu.

Drugi ogrodnik również nie stronił od nowoczesnej technologii i chętnie korzystał z pomocy zaawansowanej maszyny w swojej codziennej pracy. Prosił ją o precyzyjne prognozy suszy, wczesne wykrywanie chorób oraz analizę erozji gleby, by móc skuteczniej chronić swoje królestwo przed degradacją. Nigdy jednak nie pozwolił algorytmowi decydować o tym, które kwiaty zasługują na miano optymalnych, a które należy usunąć jako zbędny balast. Wiedział bowiem doskonale, że ogród żyje dzięki różnicy, a pewien nadmiar, asymetria i kruchość są nieodzownymi warunkami prawdziwego, nieoczywistego piękna. Maszyna pomagała mu chronić życie, ale w żadnym momencie nie próbowała definiować sensu istnienia ogrodu ani narzucać mu swojej matematycznej matrycy. Pierwszy ogrodnik zbudował bezduszną technokrację, podczas gdy drugi stworzył fundamenty Aidealizmu, czyli filozoficznej ramy będącej odpowiedzią na współczesne kryzysy prawdy, pracy i demokracji.

Współcześni krytycy koncepcji zbieżnych z Aidealizmem atakują go z kilku różnych, często sprzecznych ze sobą pozycji ideologicznych. Techno-realiści oraz sceptycy wielkich narracji ostrzegają, że tak zarysowane ramy mogą okazać się zbyt abstrakcyjne, by realnie sterować twardą polityką przemysłową państw. Ich zarzut ma swoją wagę, zwłaszcza w brutalnym świecie półprzewodników, centrów danych i narastających napięć geopolitycznych, gdzie wartości bez mechanizmów wdrożeniowych stają się jedynie poezją raportową. Odpowiedź Aidealizmu jest tu jednak jasna: filozofia ta nie ma zastępować inżynierii bezpieczeństwa ani prawa, lecz nadać im pożądany, ludzki kierunek. Kompas wszak nie jest silnikiem statku, ale nawet najpotężniejsza jednostka bez sprawnego kompasu może bardzo sprawnie i szybko dopłynąć do spektakularnej katastrofy.

Z kolei libertariańscy krytycy wszelkich regulacji alarmują, że zbyt silna kontrola nad rozwojem sztucznej inteligencji zdusi innowację i odda przewagę reżimom mniej skrupulatnym etycznie. Technologiczny libertarianizm zbyt często przypomina jednak człowieka, który wjechał windą na szczyt wieżowca, a potem ogłosił z wyższością, że schody są rozwiązaniem dla słabych. Odpowiedź Aidealizmu na te obawy jest podwójna i opiera się na głębokim zrozumieniu dynamiki społecznej. Po pierwsze, regulacja musi być inteligentna i proporcjonalna, a nie biurokratycznie ślepa na niuanse technologiczne. Po drugie, innowacja pozbawiona zaufania społecznego sama podcina swoją legitymację, gdyż technologia, której ludzie się boją lub którą uważają za narzędzie wyzysku, prędzej czy później napotka na gwałtowny opór. Etyka nie jest luksusem, na który pozwalamy sobie

po osiągnięciu wzrostu, lecz fundamentalnym warunkiem trwałości tego wzrostu w długiej perspektywie.

Krytycy z nurtów posthumanistycznych mogą z kolei uznać, że Aidealizm zbyt kurczowo trzyma się pojęcia humanistycznej godności i przestarzałego antropocentryzmu. Skoro inteligencja może przybrać formy nieludzkie, dlaczego to właśnie człowiek ma zachować centralną pozycję w nowym porządku świata? To pytanie jest bez wątpienia intelektualnie intrygujące, ale politycznie staje się skrajnie niebezpieczne, jeśli prowadzi do rozmycia odpowiedzialności wobec realnych ludzi tu i teraz. Można wszak debatować o przyszłym statusie moralnym zaawansowanych systemów syntetycznych, lecz nie wolno używać tej debaty do osłabiania praw osób, których życie już dziś zależy od decyzji algorytmicznych. Humanizm Aidealizmu nie jest wyrazem metafizycznej pychy naszego gatunku, lecz prawną zasadą ostrożności, chroniącą nieprzejrzystość egzystencjalną, czyli prawo jednostki do posiadania sfery życia niepodlegającej stałej interpretacji i punktacji przez systemy obliczeniowe.

Krytycy katastroficznego myślenia o sztucznej inteligencji słusznie ostrzegają przed paniką, która może prowadzić do niebezpiecznej centralizacji władzy w rękach największych korporacji pod pretekstem dbałości o bezpieczeństwo. Jeśli tylko najwięksi gracze rynkowi będą w stanie spełnić wyśrubowane wymogi regulacyjne, prawo stanie się de facto barierą wejścia dla mniejszej, bardziej innowacyjnej konkurencji. To ryzyko jest realne, dlatego Aidealizm musi precyzyjnie odróżniać autentyczne bezpieczeństwo od cynicznej oligopolizacji tegoż bezpieczeństwa. Konieczne są otwarte badania, publiczne audyty oraz silne wsparcie dla mniejszych podmiotów, aby etyka AI nie stała się luksusowym towarem dostępnym wyłącznie dla najbogatszych. W tym kontekście kluczowy staje się model ryzyka AI Act, czyli struktura prawna klasyfikująca systemy według stopnia zagrożenia, co pozwala na racjonalne zarządzanie innowacją w ramach demokratycznego państwa prawa.

Wreszcie krytycy wywodzący się z tradycji realistycznej w stosunkach międzynarodowych powiedzą, że pokój kantowski jest pięknym marzeniem, ale państwa zawsze będą kierować się wyłącznie interesem i siłą. Odpowiedź na ten sceptycyzm brzmi: właśnie dlatego, że świat jest brutalny, potrzebujemy silnych i sprawnych instytucji międzynarodowych. Kant nie zakładał bynajmniej, że państwa nagle staną się aniołami, lecz wierzył, że można stworzyć porządek, w którym wojna stanie się mniej opłacalna i trudniejsza do moralnego uzasadnienia. Aidealizm w świecie sztucznej inteligencji jest realistyczny właśnie wtedy, gdy nie ufa bezgranicznie ani państwom, ani korporacjom, ani modelom, lecz buduje system ich wzajemnego ograniczania. Tylko w taki sposób możemy uniknąć stanu, w którym potwór cyfrowy nie musi ryczeć, bo ma dashboard, model ryzyka i elegancki raport zgodności, pod którym ukrywa swoją niszczycielską naturę.

Z tych wszystkich rozważań wyłania się jeden, spójny projekt: nowy kontrakt społeczny dla epoki inteligencji maszynowej, który definiuje nasze wspólne zobowiązania. Jego założenia można wypowiedzieć w formie ciągłej zasady, która powinna stać się fundamentem naszej cywilizacji po automatyzacji rozumu. Sztuczna inteligencja musi być rozliczalna, ponieważ decyzja pozbawiona odpowiedzialności jest jedynie tyranią proceduralną, a algorytm, jak wiemy, nie poda się do dymisji, lecz co najwyżej otrzyma patch. Musi ona wspierać integralność intelektualną, chroniąc nas przed stanem, jakim jest niedojrzałość cyfrowa, czyli bierność, w której rezygnujemy z krytycznego myślenia na rzecz gotowych, syntetycznych odpowiedzi. Konieczne jest również dążenie do powszechnej dywidendy AI, czyli koncepcji sprawiedliwego podziału korzyści ekonomicznych płynących z automatyzacji, aby zapobiec drastycznym nierównościom społecznym.

Ten nowy kontrakt musi także przeciwdziałać zjawisku, jakim jest nierówność antropotechniczna, czyli potencjalny podział ludzkości wynikający z nierównego dostępu do sprywatyzowanych ulepszeń poznawczych. AI musi służyć rozproszonemu dobrobytowi, ponieważ automatyzacja bez redystrybucji nieuchronnie prowadzi do cyfrowego feudalizmu, w którym dane są nową pańszczyzną. Musi wspierać etyczną innowację, gdyż nowość pozbawiona dobra jest jedynie skuteczniejszym sposobem na generowanie chaosu w skali globalnej. Inteligencja maszynowa musi być również ekologicznie liczona, ponieważ nawet najwyższa sprawność obliczeniowa nie usprawiedliwia ślepoty energetycznej i niszczenia biosfery. Wreszcie, musi ona chronić wolność i godność, ponieważ człowiek nie jest surowcem predykcyjnym, a jego życie nie może zostać zredukowane do zestawu danych w modelu statystycznym.

Aidealizm nie jest więc kolejną technologiczną teorią, lecz całościową teorią cywilizacji, która staje przed wyzwaniem maszynowego intelektu. Jego głównym przeciwnikiem nie jest sama technologia, lecz nasz własny konformizm wobec jej najbardziej opłacalnych, a zarazem najbardziej destrukcyjnych zastosowań. Największym zagrożeniem nie jest to, że maszyny staną się zbyt inteligentne i przejmą nad nami kontrolę w stylu filmów science-fiction. Bardziej bezpośrednie niebezpieczeństwo polega na tym, że ludzie odpowiedzialni za ich wdrożenie pozostaną zbyt mali moralnie, zbyt chciwi instytucjonalnie i zbyt leniwi filozoficznie. W świecie, w którym AI mogłaby radykalnie wzmocnić edukację, medycynę i solidarność, tolerowanie jej redukcji do narzędzia reklamy i nadzoru byłoby cywilizacyjnym skandalem. To tak, jakby wynaleźć teleskop i używać go wyłącznie do podglądania sąsiadów; technicznie jest to możliwe, ale duchowo pozostaje głęboko żałosne.

Na końcu tych rozważań powraca do nas postać Immanuela Kanta z jego surowym, ale niosącym nadzieję przesłaniem. Wieczny pokój nie jest dla niego prognozą pogody, lecz nieustannym zadaniem do wykonania, tak jak oświecenie nie jest zamkniętą epoką, lecz ciągłą dyscypliną

rozumu. Godność ludzka nie jest ozdobą uroczystych deklaracji, lecz nieprzekraczalną granicą wszelkiej optymalizacji technicznej. Sztuczna inteligencja nie jest naszym przeznaczeniem, przed którym nie ma ucieczki, lecz najważniejszą próbą, przed jaką kiedykolwiek stanęliśmy jako gatunek. Aidealizm mówi nam, że stoimy przed wyborem: albo stworzymy instytucje zdolne podporządkować inteligencję maszynową sprawiedliwości, albo ona zostanie podporządkowana istniejącym niesprawiedliwościom, czyniąc je przerażająco wydajnymi.

Nie istnieje żadna trzecia droga w postaci neutralnej technologii, która unosiłaby się beztrosko ponad realnymi problemami tego świata. Neutralność narzędzia kończy się gwałtownie tam, gdzie zaczyna się ogromna skala jego społecznych i politycznych skutków. Jeśli AI zostanie zbudowana wyłącznie dla zysku, będzie bezlitośnie maksymalizować zysk kosztem wszystkiego innego. Jeśli dla nadzoru, będzie widzieć i analizować każdy nasz krok, odbierając nam prawo do prywatności. Jeśli jednak zostanie wpisana w instytucje prawdy, solidarności i pokoju, może stać się najpotężniejszym narzędziem korekty ludzkiej niedojrzałości, jakie kiedykolwiek stworzyliśmy. Maszyna nie ma duszy, ale może odziedziczyć nasze instytucje i jeżeli będą one skorumpowane, skorumpuje się również jej funkcja. Jeżeli jednak nasze instytucje będą oświecone, AI może pomóc nam dokończyć projekt oświecenia. Ostatnie zdanie Aidealizmu powinno więc brzmieć jak zobowiązanie moralne: nie pytajmy tylko, czy AI będzie potężniejsza od człowieka, lecz czy człowiek okaże się dość dojrzały, by nie uczynić z tej potęgi pomnika własnej małości. Barbarzyństwo z interfejsem premium nadal pozostaje barbarzyństwem, a wieczny pokój bez godności byłby tylko dobrze zarządzaną ciszą.

> Sztuczna inteligencja nie jest naszym przeznaczeniem, lecz ostatecznym sprawdzianem cywilizacyjnej dojrzałości. Pytanie brzmi, czy potrafimy zbudować instytucje silniejsze niż algorytmiczny zysk, czy też ostatecznie oddamy resztki naszej wolności w zamian za cyfrową wygodę. Barbarzyństwo z interfejsem premium nadal pozostaje barbarzyństwem, a wieczny pokój bez godności byłby tylko dobrze zarządzaną ciszą.

Inspiracje

[1] "AI Ideal Governance of AI" Niklas Lidstromer

Dr Niklas Lidströmer jest lekarzem, naukowcem i badaczem specjalizującym się w powiązaniu medycyny, sztucznej inteligencji i transformacji społecznej. Związany z Instytutem Karolinska, posiada bogate doświadczenie w zakresie innowacji medycznych opartych na sztucznej inteligencji, praktyki klinicznej i inwestycji strategicznych. Posiada tytuły doktora medycyny i magistra nauk ścisłych, a także doświadczenie kliniczne w ośmiu krajach. Dr Lidströmer jest ceniony za pracę nad modelami danych medycznych kontrolowanych przez pacjentów oraz za propagowanie etycznego zarządzania sztuczną inteligencją, którą nazywa „idealizmem". Jest redaktorem obszernego opracowania „Sztuczna inteligencja w medycynie" (Springer Nature) i autorem kilku książek zgłębiających rolę sztucznej inteligencji w opiece zdrowotnej i globalnym zarządzaniu. Jego badania koncentrują się na bezpiecznej i etycznej integracji sztucznej inteligencji z systemami opieki zdrowotnej, przy jednoczesnym zapewnieniu własności danych pacjentów i demokratycznego postępu.

Dr. Niklas Lidströmer is a physician, scientist, and researcher specializing in the intersection of medicine, artificial intelligence, and societal transformation. Affiliated with the Karolinska Institutet, he has extensive experience in AI-driven medical innovation, clinical practice, and strategic investment. He holds an MD and an MSc, and has worked clinically in eight countries. Dr. Lidströmer is recognized for his work on patient-controlled health data models and his advocacy for ethical AI governance, a concept he terms 'Aidealism.' He is the editor of the comprehensive reference work 'Artificial Intelligence in Medicine' (Springer Nature) and has authored several books exploring the role of AI in healthcare and global governance. His research focuses on integrating AI safely and ethically into healthcare systems while ensuring patient data ownership and democratic progress.

Karolinska Institutet

Literatura uzupełniająca

Książki

Literatura pokrewna (Google Scholar):

[1] **"O prawach (De legibus)"**

Marek Tuliusz Cyceron

-52 | Wydawnictwo Antyczne (wydania współczesne) | ISBN: 978-83-01-14329-9

[2] **"Katon Starszy o starości (Cato Maior de senectute)"**

Marek Tuliusz Cyceron

-44 | Wydawnictwo Antyczne (wydania współczesne) | ISBN: 978-83-04-04234-5

[3] **"The AI Ideal: AIdealism and the Governance of AI"**

Niklas Lidströmer

2026 | Morgan Kaufmann

[4] **"Artificial Intelligence in Medicine"**

Niklas Lidströmer

2022 | Springer Nature

[5] **"Deep Learning"**

Ian Goodfellow, Yoshua Bengio, Aaron Courville

2016 | MIT Press | ISBN: 978-0-262-03561-3

[6] **"Groundwork of the Metaphysics of Morals"**

Immanuel Kant

1785 | Johann Friedrich Hartknoch | ISBN: 978-0521761215

[7] **"Critique of Pure Reason"**

Immanuel Kant

1781 | Johann Friedrich Hartknoch | ISBN: 978-0521657297

[8] **"Beyond Good and Evil: Prelude to a Philosophy of the Future"**

Friedrich Nietzsche

1886 | C. G. Naumann | ISBN: 978-0-14-044923-5

[9] **"Thus Spoke Zarathustra: A Book for All and None"**

Friedrich Nietzsche

1883 | Ernst Schmeitzner | ISBN: 978-0-521-60261-7

[10] "The AI Con: How to Fight Big Tech's Hype and Create the Future We Want"

Emily M. Bender, Alex Hanna

2025 | HarperCollins | ISBN: 978-0063354319

[11] "Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax"

Emily M. Bender

2013 | Morgan & Claypool Publishers | ISBN: 978-1627051002

[12] "Data Conscience: Algorithmic Siege on our Humanity"

Brandeis Hill Marshall, Timnit Gebru

2022 | Wiley | ISBN: 978-1-119-82120-5

[13] "Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence"

Kate Crawford

2021 | Yale University Press | ISBN: 978-0-300-20957-0

[14] "Understanding the Internet: Language, Technology, Media, and Power"

Kate Crawford

2014 | Oxford University Press | ISBN: 978-0-19-933036-2

[15] "AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference"

Arvind Narayanan, Sayash Kapoor

2024 | Princeton University Press | ISBN: 978-0-691-24913-1

[16] "Fairness and Machine Learning: Limitations and Opportunities"

Solon Barocas, Moritz Hardt, Arvind Narayanan

2023 | MIT Press | ISBN: 978-0-262-04813-2

[17] "The Prince"

Niccolò Machiavelli

1532 | Antonio Blado d'Asola | ISBN: 978-0140449150

[18] "Discourses on Livy"

Niccolò Machiavelli

1531 | Bernardo Giunta | ISBN: 978-0199539253

[19] "Virtue and Terror"

Maximilien De Robespierre

2007 | Verso | ISBN: 978-1-84467-582-1

[20] "Oeuvres complètes de Maximilien Robespierre"

Maximilien De Robespierre

1910 | Ernest Leroux

[21] "The Origins of Totalitarianism"

Hannah Arendt

1951 | Harcourt, Brace | ISBN: 978-0-15-670153-2

[22] "The Human Condition"

Hannah Arendt

1958 | University of Chicago Press | ISBN: 978-0-226-02598-8

[23] "The Ethics of Artificial Intelligence"

S. Matthew Liao

2020 | Oxford University Press

[24] "Human Compatible: Artificial Intelligence and the Problem of Control"

Stuart Russell

2019 | Viking

[25] "The Ethics of AI: Power, Critique, Responsibility"

Rainer Mühlhoff

2025 | Bristol University Press

[26] "Kant Machine: Critical Philosophy After AI"

Yuk Hui

2026 | Bloomsbury Academic

[27] "The Ethics of Artificial Intelligence: A Philosophical Introduction"

Sven Nyholm

2026 | Oxford University Press

[28] "AI Ethics"

Mark Coeckelbergh

2020 | MIT Press

[29] **"Critical Theory of AI"**

Simon Lindgren

2024 | Polity

[30] **"Robespierre: A Revolutionary Life"**

Peter McPhee

2012 | Yale University Press

Artykuły naukowe

Autorzy przywoływani:

[1] **"Artificial Intelligence and the Future of Humans"**

J. Anderson, L. Rainie

2018 | Pew Research Center

Google Scholar

[2] **"The Role of Machine Learning and Artificial Intelligence in Drug Discovery and Clinical Care of Pulmonary Hypertension"**

Evelyne Bishof, Chaim Haber, Niklas Lidströmer

2026 | Pulmonary Hypertension

Google Scholar

[3] **"The association between vital signs at hospital admission and adverse outcomes in patients with COVID-19: a retrospective cohort study"**

M. Murzabekov, Niklas Lidströmer, N. Borg, M. Runold, E. Herlenius, S. Rautiainen

2025 | Frontiers in Medicine | DOI: 10.3389/fmed.2025.1602129

Google Scholar

[4] **"Ethics of artificial intelligence: Issues and initiatives"**

Luciano Floridi, Josh Cowls

2019 | Ethics and Information Technology

Google Scholar

[5] **"Generative Adversarial Nets"**

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio

2014 | Advances in Neural Information Processing Systems (NeurIPS) | DOI: 10.48550/arXiv.1406.2661

Google Scholar

[6] **"Neural Machine Translation by Jointly Learning to Align and Translate"**

Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio

2014 | International Conference on Learning Representations (ICLR) | DOI: 10.48550/arXiv.1409.0473

Google Scholar

[7] **"Why AI Ethics Is a Critical Theory"**

Waelen, Vincent

2022 | Philosophy & Technology

Google Scholar

[8] **"Kantian Deontology Meets AI Alignment: Towards Morally Grounded Fairness Metrics"**

Anonymous Researchers

2024 | arXiv

Google Scholar

[9] **"Philosophy and Ethics in the Age of Artificial Intelligence: Bridging the Gap between Technology and Human Values"**

Various Authors

2025 | Journal of Information Systems Engineering and Management

Google Scholar

[10] **"On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜"**

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, Margaret Mitchell

2021 | Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency | DOI: 10.1145/3442188.3445922

Google Scholar

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

Emily M. Bender, Alexander Koller

2020 | Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics | DOI: 10.18653/v1/2020.acl-main.463

Google Scholar

[12] **"Building the ethical AI framework of the future: from philosophy to practice"**

Various Authors

2026 | arXiv

Google Scholar

[13] **"On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?"**

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, Margaret Mitchell

2021 | Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT) | DOI: 10.1145/3442188.3445922

Google Scholar

[14] **"Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification"**

Joy Buolamwini, Timnit Gebru

2018 | Proceedings of Machine Learning Research

Google Scholar

[15] **"Principles alone cannot guarantee ethical AI"**

Brent Mittelstadt

2019 | Nature Machine Intelligence

Google Scholar

[16] **"Excavating AI: The politics of images in machine learning training sets"**

Kate Crawford, Trevor Paglen

2021 | AI & Society | DOI: 10.1007/s00146-021-01220-3

Google Scholar

[17] **"The Social Construction of Datasets"**

Will Orr, Kate Crawford

2023 | New Media & Society | DOI: 10.1177/14614448231175620

Google Scholar

[18] **"Semantics derived automatically from language corpora contain human-like biases"**

Aylin Caliskan, Joanna J. Bryson, Arvind Narayanan

2017 | Science | DOI: 10.1126/science.aal4230

Google Scholar

[19] **"AI as Normal Technology"**

Arvind Narayanan, Sayash Kapoor

2025 | Knight First Amendment Institute

Google Scholar

[20] "Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence"

Christian Zednik

2021 | Philosophy & Technology

Google Scholar

[21] "Leakage and the reproducibility crisis in ML-based science"

Sayash Kapoor, Arvind Narayanan

2023 | Patterns | DOI: 10.1016/j.patter.2023.100804

Google Scholar

[22] "The Foundation Model Transparency Index"

Rishi Bommasani, Kevin Klyman, Sayash Kapoor, et al.

2023 | Transactions on Machine Learning Research

Google Scholar

[23] "The Philosophy and Ethics of AI: Conceptual, Empirical, and Technological Investigations into Values"

Judith Simon, Gernot Rieder, Jason Branford

2024 | Digital Society

Google Scholar

[24] "Deep learning"

Yann LeCun, Yoshua Bengio, Geoffrey Hinton

2015 | Nature | DOI: 10.1038/nature14539

Google Scholar

[25] "Gradient-based learning applied to document recognition"

Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner

1998 | Proceedings of the IEEE | DOI: 10.1109/5.726791

Google Scholar

[26] "Artificial intelligence (AI) and ethical concerns: a review and research agenda"

Various Authors

2025 | AI and Ethics

Google Scholar

[27] **"Kantian deontology for AI: alignment without moral agency"**

Oluwaseun Damilola Sanwoolu

2025 | AI and Ethics

Google Scholar

[28] **"The Terror of Virtue: The French Revolution and the Problem of Political Violence"**

Marisa Linton

2017 | French Historical Studies

Google Scholar

Tekst opracowany z udziałem sztucznej inteligencji.
Redakcja i kontrola merytoryczna: Fundacja Dobre Państwo.
APA ONE Publishing System

Article ID: 500a0b4d-d3c5-42a2-bfc5-4739f450f2cd