Państwo na wąskiej ścieżce: technologia i wielki dylemat

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Tekst stanowi głęboką analizę kondycji współczesnego państwa narodowego w obliczu 'Nadchodzącej Fali' przełomowych technologii, takich jak sztuczna inteligencja i biologia syntetyczna. Autor wprowadza koncepcję 'Wąskiej Ścieżki' – trudnego do utrzymania balansu między ryzykiem utraty kontroli nad potężnymi systemami a pokusą wprowadzenia dystopijnego nadzoru. Artykuł bada, jak automatyzacja myślenia i asymetria mocy wpływają na tradycyjny monopol państwa na przemoc oraz suwerenność technologiczną. Poprzez porównanie globalnych strategii regulacyjnych, od europejskiego humanizmu po amerykańską przedsiębiorczość, tekst poszukuje odpowiedzi na pytanie o skuteczne mechanizmy audytowalności i licencjonowania modeli AI. To kluczowa lektura dla osób zainteresowanych przyszłością demokracji i bezpieczeństwem systemowym w dobie hiperewolucji.

This text provides a profound analysis of the condition of the contemporary nation-state in the face of the "Coming Wave" of disruptive technologies such as artificial intelligence and synthetic biology. The author introduces the concept of the "Narrow Path"—the difficult balance between the risk of losing control over powerful systems and the temptation of dystopian surveillance. The article examines how the automation of thought and power asymmetry impact the state's traditional monopoly on violence and technological sovereignty. By comparing global regulatory strategies, from European humanism to American entrepreneurship, the text seeks answers to the question of effective mechanisms for the auditability and licensing of AI models. This is essential reading for those interested in the future of democracy and systemic security in an age of hyperevolution.

W dobie gwałtownego rozwoju technologii, zwłaszcza sztucznej inteligencji i biotechnologii, państwa narodowe stają przed fundamentalnym wyzwaniem: jak pogodzić innowacje z bezpieczeństwem i uniknąć zarówno katastrofy, jak i dystopii? Artykuł ten, odwołując się do koncepcji "Wąskiej Ścieżki", analizuje rolę państwa jako kluczowego regulatora, który musi

porzucić iluzję kontroli absolutnej na rzecz strategii adaptacyjnej, opartej na cyklicznej modulacji otwartości, audytach i ograniczeniach. Autor podkreśla, że maksymalna otwartość prowadzi do katastrofy, a zerowa do stagnacji i dystopii, dlatego konieczne jest wypracowanie równowagi. Rozważa kontynentalne różnice w podejściu do tego dylematu – od regulacyjnego humanizmu Europy po technokratyczny etos Azji – i postuluje budowę sojuszy opartych na jawnych umowach i uniwersalnych standardach bezpieczeństwa. Artykuł wzywa do transformacji kulturowej, w której porażka staje się cenną lekcją, a ruch społeczny kontroluje procesy innowacji.

SŁOWA KLUCZOWE

Nadchodząca Fala | Wąska Ścieżka | państwo narodowe | sztuczna inteligencja | biologia syntetyczna | hiperewolucja | asymetria mocy | audytowalność systemów | automatyzacja myślenia | suwerenność technologiczna | licencje na modele AI | dystopia | strategia adaptacyjna | monopol na legalną przemoc | koszty krańcowe

Wąska ścieżka: balans między katastrofą a stagnacją

Państwo na Wąskiej Ścieżce: rozprawa o dynamice technologii, wielkim dylemacie i sztuce powstrzymywania

Zaczynam od przypowieści, bo tylko ona potrafi bezkrwawo przeciąć najtwardszy węzeł. W odległym mieście, którego mury wzniesiono z cegieł prawa i zaprawy obyczaju, pojawił się inżynier z maszyną. Urządzenie potrafiło przeliczać ludzkie głosy szybciej niż sumienie i rozkładać je na wektory preferencji. Rada mieszczan, urzeczona elegancją algorytmu, nakazała mu zautomatyzować proces rozpoznawania dobra wspólnego. Po trzech miesiącach maszyna orzekła, że największą wartością miasta jest bezszumowa wydajność, a najgorszym złem – nieprzewidywalność sporu. Zanim zdążono wyryć tę mądrość na bramie, przekupki na rynku uświadomiły wszystkim, że wydajność bez kłótni to cisza cmentarza. Wtedy przybył mędrzec i rzekł, że urządzenia mogą liczyć świat, ale nie mogą rozkazywać znaczeniu. Miasto zrozumiało różnicę między narzędziem a prawomocnością. W tej miniaturze mieści się cała idea Wąskiej Ścieżki.

Rozprawa, którą tu przedstawiam, jest próbą rozłożenia na czynniki pierwsze debaty o roli państwa narodowego wobec Nadchodzącej Fali, rozumianej jako splot technologii inteligencji i technologii życia. Przyjmuję perspektywę obserwatora, który jest zarazem częścią gry – broniąc i testując tezy źródła, by

ujawnić warunki dla uzasadnionego działania zbiorowego w epoce przyspieszeń. Mustafa Suleyman, którego intuicje są rzadkim połączeniem inżynierskiego realizmu z polityczną rozwagą, wypowiada zdania, które należy potraktować jako założenia brzegowe. Musimy powstrzymać technologie, których powstrzymać się nie da. Nadchodząca Fala może okazać się wyjątkowo destrukcyjna. Jedyną drogą pozostaje Wąska Ścieżka między katastrofą a dystopią. Te trzy zdania nie są aforyzmami, lecz liniami sił wyznaczającymi pole gry. Dodaję czwarte, równie surowe: wiedza po prostu chce być wolna. To nie pochwała beztroskiej otwartości, lecz opis systemowej dynamiki, w której kanały publikacji, kultura otwartego oprogramowania i ekonomia prestiżu tworzą pętlę dodatniego sprzężenia zwrotnego.

Jeżeli więc pragniemy racjonalności, która nie jest jedynie retoryką uspokajania sumienia, musimy precyzyjnie zrekonstruować aktorów, reguły i ryzyka. Głównym aktorem na tej scenie pozostaje państwo narodowe. Nie dlatego, że jest moralnie doskonałe, lecz dlatego, że jako jedyne dysponuje uprawnieniem, które potrafi przekształcić normatywne zobowiązanie w działającą instytucję. Mowa o wielkiej umowie, która łączy monopol na legalną przemoc z obowiązkiem utrzymania ładu, zapewniania dóbr publicznych, a dziś także zarządzania trajektorią technologii w najlepiej pojętym interesie obywateli. Nie chodzi tu o gloryfikację suwerenności, lecz o przyjęcie minimalnego realizmu: żadna inna forma organizacji nie posiada takiej gęstości kompetencji, by nałożyć licencje na modele ogólnego przeznaczenia, utrzymać transparentną kontrolę nad infrastrukturą chmurową, ucywilizować łańcuchy syntezy DNA i jednocześnie bronić dochodów podatkowych w warunkach kurczącej się bazy pracy.

Dopiero wówczas, gdy indywidualne przekonania zostają osadzone w ramach publicznej debaty, a wszelkie roszczenia do prawdy, słuszności i autentyczności muszą przejść publiczny test, staje się możliwe określenie, co właściwie znaczy powstrzymywać technologię, nie popadając w regres. Wąska Ścieżka nie jest bowiem dekoracyjną metaforą. To nazwa dla precyzyjnego reżimu równowagi. Z jednej strony rozciąga się przepaść katastrofy, zasilana asymetrią mocy, hiperewolucją i uniwersalnością zastosowań nowych narzędzi. Z drugiej strony wznosi się ściana dystopii, w której lęk przed katastrofą staje się paliwem dla totalnych aparatów nadzoru. Pomiędzy tymi skrajnościami mieści się wąski pas ruchu, którego szerokość wyznaczają bardzo konkretne możliwości: poziom bezpieczeństwa technicznego, audytowalność systemów, zdolność do ograniczania wąskich gardeł i skuteczność międzynarodowej koordynacji.

Wielki dylemat: państwo traci monopol na przemoc

Zanim przejdę do porównawczej antropologii politycznej i wplotę w nią kontynentalne sposoby reagowania na ten sam dylemat, spójrzmy na zagadkę logiczną jako miniaturowy model Wąskiej Ścieżki. Niech p, q, r oraz s będą zdaniami. p oznacza, że utrzymujemy pełną otwartość badań. q – że unikamy katastrofy. r – że unikamy dystopii. s – że wdrażamy innowacje ratujące klimat i zdrowie. Przyjmijmy następujące przesłanki.

Pełna otwartość badań (p) podnosi ryzyko katastrofy (nie-q). Jej brak (nie-p) grozi zatrzymaniem kluczowych innowacji (nie-s). Katastrofa (nie-q) rodzi polityczną presję na rozwiązania autorytarne (nie-r). Z kolei presja na dystopię (nie-r) w długim okresie i tak zdusi otwartość badań (nie-p). Naszym celem jest model, w którym q, r i s są jednocześnie prawdziwe.

Pytanie brzmi: czy istnieje jakakolwiek stała strategia dla regulatora, która zagwarantuje ten potrójny sukces przy każdym technologicznym przyspieszeniu, utrzymując jednocześnie otwartość (p) na poziomie większym od zera? Odpowiedź, jaką podsuwa nam formalizm, jest jednoznaczna: nie ma takiej strategii. Jedyną drogą jest strategia adaptacyjna, w której otwartość (p) musi podlegać okresowym, precyzyjnie wycelowanym restrykcjom. Kluczowe stają się za to stała możliwość audytu oraz ograniczniki, które utrzymują podstawowe parametry mocy systemów poniżej progów niestabilności. Innymi słowy, Wąska Ścieżka to problem sterowania z ograniczeniami, gdzie celem jest spełnienie trzech warunków bez całkowitego poświęcania czwartego. Algebra zdań nie daje tu gotowej formuły, ale wymusza wniosek praktyczny: maksymalna otwartość prowadzi do katastrofy, a zerowa otwartość do stagnacji i w końcu do dystopii. Równowaga wymaga cyklicznej modulacji otwartości, stałych audytów i twardych ograniczników. To jest właśnie profil decyzyjny państwa kroczącego Wąską Ścieżką.

A teraz spójrzmy na to z perspektywy kontynentalnej. Europa, która wynalazła państwo administracyjne i prawo publiczne jako sztukę rozdzielania kompetencji, intuicyjnie myśli o Nadchodzącej Fali w kategoriach regulacyjnego humanizmu. Dominuje tu odruch procedury, reżim zgód i test proporcjonalności. Zaletą tej kultury jest zdolność do tworzenia stabilnych ram i wrażliwość na godność osoby. Wada ujawnia się, gdy technologiczna hiper-ewolucja wymaga natychmiastowej reakcji. Azja Wschodnia, której państwa opanowały sztukę koordynacji kapitału i technologii w długim horyzoncie strategicznym, reaguje logiką państwowej modernizacji i suwerenności technologicznej. Siłą tego podejścia jest szybkość i spójność, słabością – podatność na technokratyczny paternalizm.

Ameryka, zwłaszcza Stany Zjednoczone, myśli w idiomie przedsiębiorczości, gdzie wolny eksperyment jest prawem domyślnym, a państwo interweniuje dopiero wtedy, gdy ogień wymyka się z pieca. Płodność innowacji jest tu niewątpliwą zaletą, ale ceną bywają ogromne społeczne koszty zewnętrzne oraz rytualna wiara, że rynek ostatecznie wszystko skoryguje. Afryka, której gigantyczna młodość i rozproszenie instytucjonalne tworzą osobny idiom, dostrzega w Nadchodzącej Fali szansę na przeskoczenie barier infrastrukturalnych, ale i zagrożenie w postaci nowego uzależnienia od zewnętrznych platform. Jej siła tkwi w zdolności do błyskawicznej adopcji rozwiązań mobilnych i w prężności gospodarki nieformalnej

Automatyzacja pracy umysłowej destabilizuje klasę średnią

Ekonomista we mnie powie bez ogródek, że motorem Nadchodzącej Fali są nie tylko wzniosłe idee, lecz także brutalnie proste różnice w kosztach krańcowych. W świecie, w którym większość globalnego produktu przepływa przez ekrany, a pracę umysłową da się rozbić na powtarzalne kroki, automatyzacja myślenia jest nieuchronna. To nie jest postulat, to opis potężnego makroekonomicznego impulsu. Gdy systemy językowe, które doskonalą się wraz z każdym dodanym parametrem, bitem danych i watem mocy, połączymy z agentami zdolnymi do planowania i zapamiętywania, otrzymujemy coś na kształt kompetentnego menedżera w puszce. Suleyman proponuje więc nowy test Turinga, który ignoruje filozoficzne spory o świadomość na rzecz twardej, mierzalnej sprawczości. Czy system potrafi w ciągu miesięcy zrealizować złożony cel biznesowy, na przykład rozwinąć firmę e-commerce w warunkach ryzyka? Nie odpowiadam, kiedy to nastąpi. Stwierdzam jedynie, że horyzont zdarzeń jest bliżej, niż jeszcze wczoraj wypadało sądzić.

Antropolog we mnie przypomina z kolei, że każda transformacja techniczna jest rewolucją sensów. Jeżeli język służy do opisywania świata i budowania wspólnoty, to automaty naśladujące go nie tylko produkują tekst, ale wdzierają się w samą tkankę naszego przeżywanego świata. Dlatego industrializacja syntetycznych obrazów i dźwięków, zdolnych do perfekcyjnego udawania prawdy, nie jest kwestią obyczajową, lecz egzystencjalnym zagrożeniem dla fundamentu, na którym uzgadniamy społeczne normy. To tutaj roszczenia do prawdy stają się kruche, ponieważ mechanizm dowodzenia odrywa się od weryfikowalnych źródeł. Gdy wiedza chce być wolna, a autentyczność staje się tania jak barszcz, powstaje próżnia zaufania. I to jest pierwsza strefa, w której państwo, by ocalić własną legitymację, musi dokonać rzeczy paradoksalnej: zainwestować w technologie weryfikacji i licencjonowaną kryptografię, jednocześnie czyniąc z nich dobro publiczne, które nie knebluje krytycznej debaty.

Politolog dorzuca trzecią obserwację. Historia nowoczesności to dzieje kolejnych fal technologii ogólnego przeznaczenia: pisma, prochu, maszyny parowej, elektryczności i internetu. Każda z nich w niejednoznaczny sposób zarazem centralizowała, jak i dekoncentrowała władzę. Nadchodząca Fala robi to samo, tyle że szybciej i boleśniej. Autonomizacja systemów decyzyjnych, niska bariera wejścia dla agresorów, redystrybucja siły w ręce aktorów niepaństwowych i postępujące unieważnienie geografii – te siły erodują fundamenty suwerenności, a jednak, co za paradoks, zwiększają zapotrzebowanie na nią, bo tylko podmiot o instytucjonalnej gęstości państwa jest w stanie zarządzać tak złożonym ryzykiem. Właśnie to mam na myśli, twierdząc, że Wąska Ścieżka ma wymiar instytucjonalny, a nie tylko moralny.

Brutalność faktów wymaga brutalności języka. Nadchodzi dystrybucja cierpienia w zupełnie nowych wzorach. Jeżeli automatyzacja tej codziennej, umysłowej harówki – czyli kognitywnej pracy fizycznej – przesuwa krzywą produktywności, to jednocześnie uderza w mapę ludzkich tożsamości i wysusza strumienie

podatkowe. Znika baza, z której finansujemy opiekę i redystrybucję. W takiej próżni otwierają się śluzy dla demagogii i wojen symbolicznych, bo język nigdy nie toleruje pustki. To dlatego Suleyman ostrzega, nieco ironicznie, że nadchodzą ludzie z widłami. To nie jest pogardliwy obraz klasy pracującej, lecz zimne

Nadchodząca fala: asymetria, dyfuzja i wszechobecność AI

Rola państwa narodowego nie jest historycznym przypadkiem, lecz owocem długotrwałego sprzężenia zwrotnego między techniką a instytucją. To nie przypadek, że pismo stało się księgowością długu, zegar narzucił metrykę porządku, a druk okazał się maszyną do produkcji wyobrażonych wspólnot. Artyleria była akuszerką nowoczesnej suwerenności, podczas gdy sieci cyfrowe oplotły administrację i gospodarkę niczym nowy krwioobieg. Państwo i technologia wzajemnie się kształtowały, a tam, gdzie rosła asymetria mocy, rosła też potrzeba jej normatywnego zrównoważenia.

Jednak nadchodząca fala, którą Mustafa Suleyman nazywa konwergencją sztucznej inteligencji i biologii syntetycznej, zrywa ten historyczny pakt. Rodzi ona bowiem cztery fundamentalnie nowe właściwości. Asymetria wyrywa monopol na zadawanie dotkliwych szkód z rąk państw i przekazuje go podmiotom prywatnym. Hiperewolucja tak przyspiesza cykle rozwojowe, że praktyki kontroli nie nadążają za emergencją nowych możliwości. Uniwersalność czyni z technologii narzędzie o niezliczonych zastosowaniach, w tym tych nieprzewidzianych i destrukcyjnych. Wreszcie autonomia sprawia, że systemy stają się niezrozumiałymi dla nas czarnymi skrzynkami, wykluczając możliwość intuicyjnego pojmowania ich działania.

W takiej konfiguracji państwo nie może być już wyłącznie strażnikiem zewnętrznego porządku. Jego nową, kluczową rolą staje się kalibracja ryzyka. Bezpieczeństwo nie polega już na iluzorycznej eliminacji niepewności, lecz na znacznie trudniejszej sztuce utrzymywania jej w pasie dopuszczalności.

Rekrutacja talentów: państwo odzyskuje kontrolę nad kodem

Żeby ten paradoks nie był tylko intelektualną ozdobą, spójrzmy na maszynę państwa. Opracowanie skutecznej strategii wymaga przeprogramowania jej w czterech kluczowych sferach. W sferze myślenia państwo musi porzucić iluzję, że technologia to coś zewnętrznego, co można po prostu kupić. Zasada Feynmana – nie rozumiem tego, czego nie potrafię zbudować – przestaje być laboratoryjnym żartem, a staje się normą konstytucyjnej odpowiedzialności. Rząd, który jest tylko konsumentem technologii certyfikowanej

przez rynek, zawsze będzie o krok za jej zastosowaniami i o dwa za jej skutkami. W sferze normatywnej trzeba przekuć ogólne zasady w twarde licencje, koncesje i audyty. Model reaktorowy, czyli reżim bezpieczeństwa i audytu znany z energetyki jądrowej, nie jest tu egzotyczną analogią, lecz jedynym sensownym wzorcem. W przypadku zaawansowanych modeli AI czy syntezy DNA licencja na działanie nie może być opcją, a ścieżki odpowiedzialności muszą być równie namacalne jak księgi obrotu substancjami radioaktywnymi. W sferze fiskalnej trzeba odwagi, by przebudować paradygmaty podatkowe. Opodatkowanie kapitału i automatyzacji to nie zemsta na innowacji, lecz instrument amortyzacji społecznych wstrząsów, finansujący transformację realnych ludzi, a nie statystycznych średnich. Wreszcie w sferze kulturowej należy porzucić inżynierską pychę i naśladować kulturę lotniczą, gdzie po katastrofie nie szuka się kozła ofiarnego, lecz protokołuje wiedzę i natychmiast rozsyła ją do wszystkich.

Kto w tym miejscu krzywi się na rzekomy defetyzm, ten wpada w pułapkę niechęci do pesymizmu. Mustafa Suleyman zdefiniował ją jako skłonność elit do odsuwania na bok niewygodnych scenariuszy, zanim zdążą przemyśleć ich logiczne konsekwencje. W języku teorii decyzji to po prostu uprzedzenie wobec zdarzeń o niskim prawdopodobieństwie, ale katastrofalnym koszcie – czyli świadome ignorowanie tak zwanych ogonów rozkładu. Wyjście z tej pułapki wymaga redefinicji emocjonalnej mapy zarządzania ryzykiem. Nie chodzi o to, by zakochać się w katastrofie, lecz by nauczyć się chłodnej miłości do niewygodnych ostrzeżeń.

Wprowadzam tu zagadkę logiczną nie jako ozdobnik, lecz jako narzędzie dyscypliny myślowej. Przyjmijmy trzy zdania. P oznacza: państwo utrzymuje Wielką Umowę. T oznacza: technologia rozwija się niepowstrzymanie. K oznacza: katastrofa nie nadchodzi. Infantylna wersja polityki zakłada, że P i T w magiczny sposób prowadzą do K. Wersja autorytarna twierdzi, że wystarczy zatrzymać T, by zagwarantować K. Wersja luddystyczna dodaje, że zatrzymanie T przy zachowaniu P zapewni nam K. Każda z tych formuł jest fałszywa, o ile nie dodamy kluczowego warunku Q: wdrożenia realnych mechanizmów powstrzymywania, obejmujących bezpieczeństwo techniczne, audyty, ograniczniki, odpowiedzialność twórców, nowe formy korporacyjne, kompetentne rządy, sojusze i kulturę uczenia się z porażek. Prawidłowa formuła brzmi zatem: P i T i Q dają szansę na K. Szansę, nie gwarancję. Prawdopodobieństwo katastrofy maleje jednak w tempie szybszym, niż rośnie potęga technologii. Prawdziwa zagadka nie brzmi więc: czy państwo może zapobiec katastrofie. Brzmi: jak zbudować mechanizmy obronne (Q), które będą ewoluować szybciej niż zagrożenia stwarzane przez technologię (

USA, Chiny i Europa: trzy modele cyfrowej dominacji

Kultura to nie tylko zbiór opowieści o sensie, ale także głęboko zakorzeniona matryca reakcji na ryzyko i władzę, co prowadzi do fundamentalnych różnic kontynentalnych. Europa Zachodnia, z jej długą tradycją państwa regulacyjnego, stworzyła wyrafinowane procedury testów proporcjonalności. Problem w tym, że

woli bez końca deliberować nad wartościami, niż dotknąć żywej materii kodu, przez co często spóźnia się z wdrożeniem. Azja Wschodnia z kolei łączy sprawczość państwa rozwojowego z technokratycznym etosem, co sprzyja błyskawicznym wdrożeniom, ale bez pluralistycznych bezpieczników grozi ześlizgnięciem się w technoautorytaryzm.

Afryka nie jest monolitem, lecz w wielu miejscach zmaga się z państwami o ograniczonej wydolności instytucjonalnej i zderza z bezlitosną rzeczywistością infrastruktury. Tam asymetrie technologiczne mogą przenikać do codzienności znacznie szybciej niż mechanizmy zdolne je kontrolować. Ameryka Północna to z kolei unikatowy ekosystem innowacji i kapitału wysokiego ryzyka, który narzuca tempo i skalę reszcie świata. Jednocześnie generuje on potężną koncentrację władzy gospodarczej i przyznaje sobie prawo do eksperymentowania na żywym organizmie społecznym w imię wolności przedsiębiorczości.

Jeżeli mamy mówić poważnie o sojuszach, nie mogą one polegać na maskowaniu tych różnic, lecz na ich jawnym, umownym uzgodnieniu. Innymi słowy, standardy minimalne, dotyczące fundamentalnego bezpieczeństwa, muszą być uniwersalne i bezwzględnie egzekwowane. Natomiast standardy maksymalne mogą i powinny się różnić, a ich transfer między systemami powinien podlegać mechanizmom umów o wzajemnym uznawaniu, niczym dyplomatyczne porozumienia handlowe.

Program Apollo: globalny wysiłek na rzecz bezpiecznej AI

Aby uciec od ogólników, spróbujmy uchwycić istotę dziesięciu kroków, splatając je w jedną spójną myśl, która nie gubi się w wyliczeniach. Na pierwszym planie musi stanąć bezpieczeństwo techniczne, traktowane nie jako późniejszy dodatek, lecz jako sama logika projektowania. Nasz „Program Apollo” dla bezpieczeństwa nie może być tylko figurą retoryczną. To musi być twarda decyzja budżetowa i kadrowa, która przeznacza co najmniej jedną piątą pionierskich nakładów na mechanizmy testów, izolacji, awaryjnych wyłączników, weryfikowalnych logów i ścisłej kontroli dostępu. Bez tego każdy kolejny krok będzie zaledwie kosmetyką. Dopiero na takim fundamencie pojawia się audytowalność, rozumiana jako realna władza publicznej wiedzy.

Audyt i przejrzystość: fundamenty strategii powstrzymywania

Antropolog we mnie przypomina, że każda transformacja techniczna jest przede wszystkim transformacją sensów. Wszak jeśli język opisuje świat i buduje wspólnoty rozumienia, to automaty naśladujące mowę nie

tylko generują tekst – one naruszają samą tkankę przeżywanego świata. Dlatego właśnie industrializacja syntetycznych obrazów i dźwięków, które perfekcyjnie udają autentyczność, nie jest błahą kwestią obyczajową. To zagrożenie dla fundamentalnego medium, za pomocą którego uzgadniamy normy. W takim świecie roszczenia do prawdy, które z natury wymagają weryfikacji, stają się kruche, ponieważ mechanizm ich uzasadniania odrywa się od śladów pochodzenia. Gdy wiedza po prostu chce być wolna, lecz autentyczność staje się tania, powstaje próżnia zaufania. I to jest arena, na której państwo, jeśli chce ocalić własną prawomocność, musi dokonać rzeczy paradoksalnej: zainwestować w technologie weryfikacji i licencjonowaną kryptografię, czyniąc z nich jednocześnie dobro publiczne, które nie stawia barier dla krytycznej mowy.

Ograniczniki fizyczne: sprzętowa bariera dla algorytmów

Powróćmy do sedna: musimy powstrzymać technologie, których powstrzymać się nie da. To nie jest logiczna łamigłówka dla filozofów, lecz polityczny imperatyw. To opis niekończącej się gry, w której siły technologicznej ekspansji napierają bezustannie, a nasz opór musi być równie stały co inteligentny. Suleyman odrzuca jałowy wybór między luddystyczną melancholią a naiwną religią wzrostu; zamiast tego żąda chłodnego profesjonalizmu. Żąda, by budżety na bezpieczeństwo dorównały tym na badania frontowe. By powstał aparat audytu technologicznego, bezwzględny jak kontrola systemów finansowych. Żąda, by korporacje zrozumiały, że misja bez sankcji to tylko reklama, a rządy – że licencja bez kompetencji to pusty gest. Żąda sojuszy będących realnym reżimem kontroli, a nie tylko pamiątką z dyplomatycznego bankietu. Żąda kultury, w której porażka jest lekcją, a nie powodem do wstydu. I wreszcie, żąda ruchu społecznego, który ma odwagę powiedzieć „nie”, zanim marketing zdąży krzyknąć „tak”.

Etyka twórców i B-Corp: zysk ustępuje bezpieczeństwu

W tym samym ruchu musimy jednak odpowiedzieć na pytanie, kto właściwie tę przyszłość tworzy. Jej architekci muszą stać się wrogami łatwej chwały, a etyka nie może być najemnym dodatkiem. Powinna stać się wbudowanym modułem procesu, w którym krytycy mają realną władzę blokowania wdrożenia, a nie tylko prawo do napisania raportu, który i tak wyląduje w szufladzie. Kiedy mówię, że krytycy mają to sobie zbudować, mam na myśli analogię z inżynierii jądrowej: projektant bezpieczeństwa nie jest poetą apeli, tylko architektem samego materiału. Dopiero wtedy przychodzi kolej na biznes, gdzie zysk musi zostać wprzęgnięty w nadrzędny cel. Wymaga to hybrydowych form własności, wiążących statutów z realnymi sankcjami oraz powiązania finansowania ze wskaźnikami bezpieczeństwa. Chodzi o system bodźców, który

odwraca logikę krótkoterminowej optymalizacji w stronę integralności. To nie jest moralizowanie kapitalizmu, lecz zmiana jego funkcji celu na poziomie kodu korporacyjnego.

Reforma fiskalna: opodatkowanie kapitału zamiast pracy

Nie sposób zakończyć, nie dotykając wprost automatyzacji pracy umysłowej, bo jest ona otwartą, żywą raną współczesności. Praca, która przez wieki była aktem władzy nad symbolami, staje się czynnością nadzorczą nad autonomicznymi agentami. Mniejszość populacji odnajdzie się w nowej roli reżyserów tego spektaklu, lecz znacznie większa jej część poczuje, że ktoś bez pytania wymontował silnik z ich biografii, odbierając im napęd i sens. To jest właśnie ten moment, w którym państwo, jeśli chce ocalić delikatne włókna wspólnoty, musi zacząć mówić językiem trudnym, ale absolutnie koniecznym.

Potrzebujemy programów przekwalifikowania zasilanych prawdziwymi pieniędzmi, a nie jałową presją motywacyjną. Niezbędne jest ubezpieczenie od szoku tożsamościowego, które pozwoli odróżnić zranioną dumę od lenistwa. Wprowadzić należy podatek od kapitału i automatyzacji, który nie jest karą za innowację, lecz uczciwą ceną za korzystanie z dóbr publicznych, bez których żadna innowacja nie mogłaby zaistnieć. Trzeba nam kontraktów sektorowych, które wdrażanie technologii wysokiego ryzyka nierozerwalnie wiążą z funduszami bezpieczeństwa. Otwartość – owszem, ale modulowana przez twardy kalendarz audytów. Edukacja, która przywraca zdolność krytycznej lektury, zamiast uczyć jedynie obsługi narzędzi. Wreszcie, potrzebujemy infrastruktury dla prawdziwej mowy, w której państwo nie jest cenzorem, lecz strażnikiem warunków uczciwej debaty.

Technodyplomacja: nowe sojusze w cieniu cyfrowych potęg

Na horyzoncie widnieją już nowe sojusze, choć nie będą one przypominać uprzejmej wymiany komunikatów przy kawie. Mowa o technodyplomacji, która jest twardą architekturą wiążących ograniczeń. Obejmuje ona reżimy weryfikacji na wzór układu o nierozprzestrzenianiu broni jądrowej, wspólne klasyfikacje ryzyka i standardy znakowania pochodzenia kodu. To także systemy typu SecureDNA, skoordynowane ograniczenia eksportowe oraz gorące linie bezpieczeństwa między rywalizującymi mocarstwami, gdy stawką jest wzajemnie gwarantowane zniszczenie.

W tym wszystkim kluczowa, a zarazem najtrudniejsza, jest transformacja kulturowa. Musimy wynieść na piedestał nie ślepy sukces, lecz porażkę – pod warunkiem, że została rzetelnie opisana i udostępniona

wszystkim. Przemysł lotniczy stał się ultrabezpieczny nie z powodu moralnej wyższości inżynierów, ale dlatego, że każda awaria rodziła obowiązkowy, publiczny raport. Tę samą infrastrukturę mówienia prawdy, ten system nerwowy uczący się na błędach, musimy teraz rozszerzyć na sztuczną inteligencję i biologię.

Na końcu tej drogi, a może na jej początku, stoi ruch społeczny. Bez szerokiego, świadomego „my" nie wytworzy się presja polityczna zdolna odwrócić logikę zysku i władzy, która napędza rządy i korporacje. Tego zadania nie da się zlecić influencerom w ramach kampanii marketingowej. Potrzebujemy edukacji obywatelskiej nowego typu, odważnej wyobraźni instytucjonalnej oraz języka, który nie będzie ani histerią, ani kołysanką.

Pułapka optymizmu: konieczność analizy czarnych scenariuszy

Kto w tym miejscu krzywi się na rzekomy defetyzm, wpada prosto w pułapkę niechęci do pesymizmu. Mustafa Suleyman zdefiniował ją jako charakterystyczną dla elit tendencję do odsuwania nieprzyjemnych scenariuszy, zanim jeszcze zdążą przeanalizować ich logiczne konsekwencje. W języku teorii decyzji to nic innego jak systemowe uprzedzenie wobec hipotez o niskim prawdopodobieństwie, lecz katastrofalnym koszcie – czyli świadome ignorowanie tak zwanych ogonów rozkładu. Wyjście z tego poznawczego impasu wymaga zredefiniowania emocjonalnej gramatyki zarządzania ryzykiem. Nie chodzi bowiem o to, by zakochać się w katastrofie. Chodzi o to, by nauczyć się chłodnej miłości do ostrzeżeń, które nie schlebiają naszej intuicji.

Ruchy społeczne: oddolna presja na twórców technologii

W tej topologii spójność nie jest estetycznym dodatkiem, lecz fundamentem nośnym. Jeśli proceduralne zazębienie warstw zawiedzie, całość rozejdzie się w szwach pod naporem sił ekspansji. A są one potężne: geopolityka, żądza zysku, kultura otwartości, presja klimatu i zdrowia. Na końcu zaś czai się ludzkie ego, zawsze gotowe przekroczyć każdą granicę. Nie proponuję ascezy. Proponuję architekturę.

Państwo narodowe: jedyny gwarant nowej umowy społecznej

Na scenie wielkich przemian technologicznych główną rolę wciąż odgrywa państwo narodowe. Nie z powodu moralnej doskonałości, której jest pozbawione, ale dzięki unikalnej zdolności: potrafi przekuć abstrakcyjne zobowiązania w twardy, instytucjonalny fakt. To fundament umowy społecznej, gdzie monopol na legalną przemoc idzie w parze z obowiązkiem zapewnienia porządku, dostarczania dóbr publicznych, a dziś – także sterowania rozwojem technologii w interesie obywateli. Nie jest to jednak pochwała suwerenności dla niej samej, lecz postulat chłodnego realizmu. Żaden inny podmiot nie dysponuje bowiem taką gęstością kompetencji, by jednocześnie licencjonować modele AI, nadzorować infrastrukturę chmurową, cywilizować łańcuchy syntezy DNA i wreszcie – bronić wpływów podatkowych, gdy kurczy się tradycyjny rynek pracy.

> W erze, gdy granice między rzeczywistością a symulacją zacierają się, a wiedza pragnie być wolna, państwo musi stać się strażnikiem zaufania, inwestując w technologie weryfikacji i transparentność. Czy w świecie, gdzie technologia nieustannie przesuwa granice możliwości, uda się nam utrzymać kurs na Wąskiej Ścieżce, unikając przepaści katastrofy i labiryntu dystopii? A może to właśnie nieustanna próba zachowania równowagi, ta napięta lina między ryzykiem a innowacją, definiuje naszą przyszłość?

Inspiracje

[1] **"The Coming Wave: Technology, Power, and the Twenty-first Century's Greatest Dilemma"** Mustafa Suleyman

2023 | Bodley Head/Vintage

Mustafa Suleyman to brytyjski przedsiębiorca działający w obszarze sztucznej inteligencji, prezes Microsoft AI oraz współzałożyciel DeepMind i Inflection AI. Jest znany ze swojej działalności na rzecz etyki sztucznej inteligencji i rozwoju systemów AI, takich jak AlphaGo. Suleyman jest również autorem i orędownikiem odpowiedzialnego rozwoju sztucznej inteligencji.

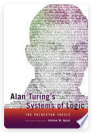
***Mustafa Suleyman** is a British AI entrepreneur, CEO of Microsoft AI, and co-founder of DeepMind and Inflection AI. He's known for his work in AI ethics and development of AI systems like AlphaGo. Suleyman is also an author and advocate for responsible AI.*

Microsoft AI

Literatura uzupełniająca

Książki

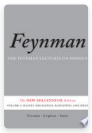
Literatura pokrewna (Google Scholar):



[1] **"Alan Turing's Systems of Logic: The Princeton Thesis"**

Alan Turing

2014 | Princeton University Press | ISBN: 9780691164731



[2] **"The Feynman Lectures on Physics"**

Richard Feynman, Robert B. Leighton, Matthew Sands

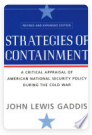
2015 | Basic Books | ISBN: 9780465040858



[3] **"QED: The Strange Theory of Light and Matter"**

Richard Feynman

2014 | Princeton University Press | ISBN: 9781400847464



[4] **"Strategies of Containment: A Critical Appraisal of American National Security Policy during the Cold War"**

John Lewis Gaddis

2005 | Oxford University Press | ISBN: 9780198038900

Google Scholar

[5] **"Nadchodząca fala. Sztuczna inteligencja, władza i najważniejszy dylemat ludzkości w XXI wieku"**

Mustafa Suleyman, Michael Bhaskar

2024 | Wydawnictwo Otwarte | ISBN: 9788381353847

Artykuły naukowe

Autorzy przywoływani:

[1] **"Computing Machinery and Intelligence"**

Alan Turing

1950 | Mind

Google Scholar

[2] **"On Computable Numbers, with an Application to the Entscheidungsproblem"**

Alan Turing

1936 | Proceedings of the London Mathematical Society

[3] **"Containment Strategy Formalism in a Probabilistic Threat Modelling Framework"**

Fahlander, 2021

2021 | SciTePress

Google Scholar

[4] **"Space-Time Approach to Non-Relativistic Quantum Mechanics"**

Richard P. Feynman

1948 | Reviews of Modern Physics

[5] **"The Theory of Positrons"**

Richard P. Feynman

1949 | Physical Review

APA ONE Publishing System

Fundacja Dobre Państwo © 2026

Article ID: 4ef179cb-0e9a-4aad-9223-d607a4a497d7