Nadchodząca fala: technologia, państwo i architektura powstrzymywania

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje nadchodzącą rewolucję technologiczną, określaną mianem Nadchodzącej Fali, która łączy potencjał sztucznej inteligencji oraz biologii syntetycznej. Autor bada, w jaki sposób te potężne narzędzia wpływają na fundamenty państwa narodowego, wymuszając stworzenie nowej architektury powstrzymywania. Tekst przybliża koncepcję Wąskiej Ścieżki – strategii balansu między innowacją a bezpieczeństwem. Poruszane są kluczowe kwestie, takie jak problem goryli, automatyzacja pracy umysłowej oraz konieczność audytów technologicznych i kontroli łańcuchów dostaw półprzewodników. Analiza wskazuje na potrzebę redystrybucji zysków z AI oraz wzmocnienia suwerenności technicznej w obliczu systemowych zmian, które rzucają wyzwanie tradycyjnej legitymacji instytucji publicznych i stabilności społeczno-ekonomicznej.

This article analyzes the coming technological revolution, dubbed the Coming Wave, which combines the potential of artificial intelligence and synthetic biology. The author examines how these powerful tools are impacting the foundations of the nation-state, forcing the creation of a new architecture of containment. The paper introduces the concept of the Narrow Path—a strategy for balancing innovation and security. It addresses key issues such as the gorilla problem, the automation of white-collar work, and the need for technological audits and control of semiconductor supply chains. The analysis points to the need to redistribute AI gains and strengthen technological sovereignty in the face of systemic changes that challenge the traditional legitimacy of public institutions and socioeconomic stability.

Artykuł analizuje wyzwania związane z nadchodzącą falą technologiczną, definiowaną jako sojusz sztucznej inteligencji i biologii syntetycznej, która charakteryzuje się asymetrią, wszechstronnością, autonomią i wykładniczym tempem rozwoju. Autor, opierając się na koncepcjach Mustafy Suleymana, argumentuje, że beztroska otwartość prowadzi do katastrofy, paniczna kontrola do dystopii, a rezygnacja z inżynieryjnego rozumu do stagnacji.

Rozwiązaniem jest „Wąska Ścieżka” – sztuka równowagi między innowacją a bezpieczeństwem, wymagająca ograniczonej otwartości i kontroli, audytów oraz strategicznych wąskich gardeł, takich jak licencjonowanie niebezpiecznych kompetencji. Kluczowe jest również ponowne zdefiniowanie roli państwa w kontekście automatyzacji, w tym redystrybucja zysków z inteligencji i tworzenie mechanizmów fiskalnych, aby uniknąć masowej utraty godności i legitymacji. Artykuł podkreśla potrzebę międzynarodowej współpracy, odpowiedzialności twórców i zmiany kultury innowacji, aby zapobiec katastrofie.

SŁOWA KLUCZOWE

Nadchodząca fala | biologia syntetyczna | architektura powstrzymywania | problem goryli | wąska ścieżka | automatyzacja pracy umysłowej | audyt technologiczny | łańcuch dostaw półprzewodników | redystrybucja zysków z AI | legitymacja państwa | suwerenność techniczna | systemy autonomiczne | wąskie gardła | kognitywna praca fizyczna | technodyplomacja

Nadchodząca fala: wszechobecność i niska bariera wejścia

Aby zrozumieć teraźniejszość, trzeba uchwycić rytm historii: technologie ogólnego przeznaczenia nadchodzą falami, a każda fala rzeźbi na nowo instytucje. Ewolucja naszej mocy to historia manipulacji – najpierw twardą materią, potem abstrakcyjnymi znakami, a dziś samymi modelami inteligencji i kodem życia. Gdy Mustafa Suleyman definiuje nadchodzącą falę jako sojusz sztucznej inteligencji i biologii syntetycznej, stawia diagnozę totalną. Nie jest to spekulacja, lecz opis systemowego sprzężenia, w którym zdolność do działania staje się radykalnie tańsza, a jej siła rozkłada się asymetrycznie. Właśnie dlatego hiperewolucja, wszechstronność i autonomia przestają być figurami retorycznymi. Stają się nazwami wektorów, które wyznaczają nową architekturę możliwości.

Problem goryli: ryzyko utraty kontroli nad nadinteligencją

Kluczowy staje się tu problem goryli, metafora zapożyczona przez Suleymana od Stuarta Russella, przy czym nie jest to analogia z zoologii, lecz z socjologii inteligencji jako zasobu dominacji. Silniejszy nie

dominuje nad mądrzejszym, bowiem siła jest jednowymiarowa, a inteligencja – transformatywna. Jeśli zatem systemy planujące i uczące się osiągną kompetencję działania w świecie, wykraczając poza rekombinację symboli, to w sposób konieczny powstaje nowy rodzaj ryzyka strukturalnego. Ryzyka, którego nie amortyzuje sama dobra wola twórców. W tym punkcie wymóg ograniczników, audytów i technicznych wyłączników awaryjnych przestaje być stanowiskiem ideologicznym. Staje się wymogiem racjonalności praktycznej.

Wielki dylemat: katastrofa, dystopia lub powstrzymanie

Wielki Dylemat rozpościera się między trzema niepożądanymi skrajnościami a jednym niestabilnym centrum. Beztroska otwartość prowadzi wprost do katastrofy, paniczna kontrola rodzi dystopię, a rezygnacja z rozumu inżynieryjnego skutkuje stagnacją. Pomiędzy nimi biegnie Wąska Ścieżka, która nie jest kompromisem, tylko wymagającą sztuką równowagi. W tej sztuce każda warstwa porządku, technicznego czy prawnego, działa dwojako: jest zarazem ograniczeniem i źródłem nowych możliwości. Program ten może brzmieć jak asceza, lecz w istocie jest sztuką nadawania formy.

Dyfuzja wiedzy: nieuchronny przepływ informacji

Wiedza po prostu chce być wolna. To stwierdzenie należy traktować nie jako slogan, lecz jako prawo termodynamiki informacji. Cały ekosystem – od finansowania badań, przez systemy reputacji naukowej, po repozytoria preprintów i otwarte oprogramowanie – tworzy potężny gradient dyfuzji, którego nie odwróci żaden edykt. Skoro tak, właściwe pytanie nie brzmi, jak zatrzymać ten proces, lecz jak zorganizować go wokół strategicznych wąskich gardeł. Chodzi o to, by dostęp do najbardziej niebezpiecznych kompetencji wymagał licencji, uwiarygodnionej tożsamości i podlegających audytowi zapisów aktywności, uniemożliwiając powstawanie dzikich laboratoriów. Suleyman proponuje właśnie takie przesunięcie: od bezradnego zakazu w stronę architektury ograniczeń, która wykorzystuje naturalne punkty koncentracji, takie jak łańcuchy dostaw półprzewodników, reglamentowany dostęp do chmury czy kontrolę nad syntezą DNA.

Wąska ścieżka: balans między wolnością a bezpieczeństwem

Parabola o wąskiej ścieżce w górach. Karawana postępuje wzdłuż ostrej grani, gdzie po lewej stronie rozciąga się przepaść katastrofy, a po prawej urwisko dystopii. Przewodnik dzierży kij utrzymujący w napięciu dwie siły: prawo do ryzyka innowacji oraz imperatyw bezpieczeństwa publicznego. Jeśli napięcie

którejkolwiek z nich osłabnie, karawana nieuchronnie traci równowagę. Największą trudność sprawia fakt, że wiatr zmienia się szybciej niż kroki. Dlatego żaden program powstrzymywania nie jest rozwiązaniem ustalonym raz na zawsze. To nieustanne ćwiczenie w oddychaniu instytucjami.

Automatyzacja kognitywna: erozja bazy podatkowej państwa

Automatyzacja pracy umysłowej wprowadza nas w nową, niepokojącą dialektykę uznania. Gdy wielkie modele językowe i zalążkowa sztuczna inteligencja przejmują zadania oparte na języku, które dotąd tworzyły PKB, znikają całe połacie tego, co można nazwać kognitywną pracą fizyczną. Wraz z nimi znikają nie tylko pensje, ale również tożsamości zawodowe, w których ludzie przez lata zakorzeniali swoje poczucie wartości. Mustafa Suleyman ma rację, gdy ostrzega, że ludzie z widłami to nie jest tania metafora populizmu, lecz przewidywalna reakcja społeczeństw na masową utratę godności i uznania. Rozsądną odpowiedzią nie jest zakaz generowania e-maili przez AI, lecz głęboka, fiskalna i edukacyjna przebudowa systemu. Chodzi o to, by zysk z inteligencji przestał być prywatnym strumieniem płynącym do właścicieli modeli, a stał się źródłem redystrybucji i finansowania usług publicznych w momencie, gdy dochody z pracy zaczynają maleć.

Spróbujmy uporządkować to rozumowanie rygorem rachunku zdań. Niech p oznacza postulat nieograniczonej otwartości, q – bezpieczeństwo publiczne, a r – demokratyczną legitymację. Jeśli przyjmiemy, że p implikuje nie-q (otwartość zagraża bezpieczeństwu), a nie-q implikuje nie-r (brak bezpieczeństwa podważa legitymację), to przez podwójne zastosowanie reguły modus ponens dochodzimy do wniosku, że p prowadzi do nie-r. Jeśli jednak założymy, że również nie-p (zamknięcie) prowadzi do nie-r, otrzymujemy formułę, w której zarówno p, jak i nie-p skutkują utratą legitymacji. To logiczny dowód na fatalizm. Aby go uniknąć, musimy wprowadzić dodatkowe aksjomaty: ograniczoną otwartość s oraz ograniczoną kontrolę t. Wtedy możemy przyjąć, że p w połączeniu z t implikuje q, a s w połączeniu z nie-t implikuje r, ale tylko pod warunkiem istnienia audytu a. Wówczas jednoczesne spełnienie q i r staje się możliwe, o ile znajdziemy taką konfigurację p, s, t, a, która minimalizuje koszt społeczny. Ten szkic dowodzi, że ani skrajna otwartość, ani jej brak nie zapewnią nam legitymacji. Wymaga ona spójnego układu a, s, t, które są technicznymi i proceduralnymi wcieleniami wąskiej ścieżki.

Automatyzacja nieuchronnie redefiniuje państwo i finanse publiczne. Gdy znika tradycyjna baza podatku od pracy, a zyski z inteligencji kumulują się w rękach właścicieli infrastruktury obliczeniowej, w budżecie powstaje luka. Luka ta zniszczy usługi publiczne, chyba że zostanie wypełniona opodatkowaniem kapitału, robotów i zysków z automatyzacji oraz opłatami licencyjnymi za dostęp do wspólnych zasobów danych. Należy bowiem pamiętać, że redystrybucja bez produktywności jest iluzją, a produktywność bez redystrybucji jest dynamitem. Zasadą musi być ich sprzężenie. Państwo, które nie zbuduje wiarygodnych i

apolitycznych mechanizmów fiskalnych dla ery ACI, skaże się na chaotyczną politykę gaszenia pożarów. Z kolei państwo, które oprze się wyłącznie na podatku od robotów, zamrozi innowacje, wpadając w pułapkę przeciętności. Potrzebny jest zdywersyfikowany koszyk instrumentów, który stabilizuje oczekiwania i finansuje transformację, zamiast jedynie tapetować rzeczywistość doraźnymi dotacjami.

USA vs. Chiny i Europa: geopolityka regulacji AI

Superfala uderza w każdy kontynent inaczej, aktywując odmienne zasoby cywilizacyjne. Azja myśli w skali dekad, integrując robotykę z biotechnologią w imię suwerenności technicznej, co nieuchronnie wzmacnia centralizację i prowadzi wprost do technomerkantylizmu. Afryka z kolei dokonuje skoku dzięki technologiom mobilnym i rozproszonej energetyce, stając się globalnym laboratorium rozwiązań niskokosztowych, ale zarazem poligonem dla firm eksternalizujących koszty swoich eksperymentów. Ameryka zaś postrzega innowację jako moralny dramat jednostkowego geniuszu, nagradzanego przez kapitał wysokiego ryzyka, co sprzyja nieokiełznanej wierze w skalowalność, która zderza się z interwencją antymonopolową, gdy jest już za późno. Europa natomiast nadaje rzeczywistości normatywne ramy, zanim sama ją zbuduje – bywa to cnotą chroniącą prawa, a bywa pustym rytuałem, gdy brakuje potencjału technologicznego.

Dekalog powstrzymywania: plan kontroli technologii

Program dziesięciu koncentrycznych kręgów powstrzymywania musi stać się spójną, zintegrowaną praktyką, a nie zbiorem luźnych zaleceń. Bezpieczeństwo, rozumiane z powagą godną Programu Apollo, wymaga, aby co najmniej jedna piąta budżetów na badania przełomowe była alokowana na prewencję ex ante, czyli analizę ryzyka przed jego wystąpieniem, a nie na heroiczne, lecz spóźnione akcje ratunkowe ex post. Audyty muszą przekształcić się w proaktywne testowanie odporności na „czarne łabędzie” – zdarzenia skrajnie nieprawdopodobne i katastrofalne w skutkach – zamiast pozostawać koleżeńskim panelem przy kawie i prezentacji. Ograniczniki rozwoju technologii powinny celować w fizyczne wąskie gardła: dostęp do materiałów, energii i unikalnych kompetencji, ale nie mogą stać się protekcjonizmem w przebraniu.

Twórcy muszą mieć wbudowany prawny obowiązek odpowiedzialności, co oznacza, że rekurencyjne samodoskonalenie systemu, które znosi ludzką kontrolę, jest sygnałem alarmowym, a nie trofeum do zdobycia. Biznes z kolei powinien porzucić fikcję, że standardowy model korporacyjny jest w stanie realizować cele społeczne, i zacząć wdrażać konstrukcje hybrydowe, które trwale wiążą wynik finansowy z weryfikowalnymi wskaźnikami bezpieczeństwa. Rządy muszą odzyskać zdolność do tworzenia własnej

technologii, ponieważ regulacja bez rozumienia jest magią, a magia w polityce państwowej zawsze kończy się bolesnym procesem przed trybunałem rzeczywistości.

Sojusze międzynarodowe należy budować nie na fundamencie moralizatorskiej retoryki, lecz w twardym języku weryfikowalnych warunków, na wzór Protokołu Montrealskiego, z klauzulami snapback, czyli mechanizmami natychmiastowego przywracania sankcji, oraz wspólnymi platformami testowymi. Kultura innowacji powinna nagradzać ujawnianie porażek, jak czyni to branża lotnicza, a nie karać za nie, jak nakazuje PR; w przeciwnym razie incydenty będą ukrywane aż do nieuniknionej katastrofy. Wreszcie ruchy społeczne muszą przezwyciężyć polaryzację i zdobyć kompetencje technologiczne, gdyż gniew pozbawiony zrozumienia produkuje jedynie efemeryczne zwycięstwa memów. Spójność wszystkich tych warstw jest warunkiem powodzenia całego programu, ponieważ każda z nich, działając w izolacji, jest jak pas bezpieczeństwa bez samochodu.

Państwo-programista: budowa suwerennych technologii

Państwo narodowe nie jest anachronizmem, lecz jedynym aktorem zdolnym do podjęcia niewdzięcznej pracy systemowej: od licencji i standardów, przez konfiskatę zysków z efektów zewnętrznych, po technodyplomację. Współczesna Wielka Umowa jest transakcją ostateczną: oddajemy monopol na użycie siły w zamian za monopol na odpowiedzialność. Jeśli państwo zawiedzie w epoce syntetycznych mediów i taniej agresji, imploduje samo źródło jego legitymacji.

Modele ogólne: ryzyko autonomicznego działania systemów

Analiza ta dekonstruuje fundamentalne napięcie współczesności, w którym państwo narodowe – wciąż jedyny podmiot zdolny utrzymać ład i odpowiedzialność instytucjonalną – zderza się z hiperewolucyjną dynamiką sztucznej inteligencji i biologii syntetycznej. To właśnie tę siłę Mustafa Suleyman określił mianem Nadchodzącej Fali. Jej moc wynika z czterech dystynktywnych cech: asymetrii, wszechstronności, autonomii i wykładniczego tempa rozwoju. To one destabilizują ekonomiczne i polityczne zdolności państw, asymetrycznie wzmacniając aktorów niepaństwowych i w konsekwencji redefiniując fundamentalną strukturę władzy.

Wąskie gardła: fizyczne limity produkcji procesorów

Właściwe pytanie nie dotyczy zatem tego, czy ograniczać, lecz jak zorganizować dyfuzję technologii wokół jej naturalnych wąskich gardeł. Chodzi o system, w którym dostęp do najbardziej niebezpiecznych kompetencji wymaga licencji, uwiarygodnionej tożsamości i ścieżki audytu, co uniemożliwi powstawanie dzikich laboratoriów i czarnych skrzynek nieznanego pochodzenia. Mustafa Suleyman postuluje tu fundamentalne przesunięcie: od bezradnego zakazu ku inteligentnej architekturze ograniczeń. Architekturze, która opiera się na istniejących punktach koncentracji mocy, takich jak łańcuchy dostaw półprzewodników, bramkowany dostęp do chmur obliczeniowych czy kontrolowana synteza DNA.

Dywidenda z AI: redystrybucja jako fundament stabilności

Automatyzacja pracy umysłowej nie jest epizodem, lecz trwałym wektorem przesuwającym wartość z pracy w stronę czystego kapitału. Maszyny uczą się pisać, podczas gdy budżety uczą się obumierać, jeśli nie zredefiniują swoich źródeł dochodu. Mustafa Suleyman nazwał to wprost: w długim okresie AI jest narzędziem fundamentalnie zastępującym pracę. Jeśli nie przebudujemy krwiobiegów fiskalnych, organizm państwa będzie oddychać coraz płycej.

Audyty i red teaming: weryfikacja bezpieczeństwa modeli

Fraktal odpowiedzialności ma najmniejsze ogniwo w kodzie, gdzie zaczyna się łańcuch zabezpieczeń: od bezpiecznych bibliotek po ograniczenia w samych modelach. Wyżej w tej strukturze stoją audyty, zespoły typu red team, skrupulatne rejestry aktywności i automaty testujące inne automaty. Następnie pojawiają się ograniczenia w łańcuchach dostaw oraz licencje na moc obliczeniową, syntezę i dostęp do chmury. Ponad nimi jest kultura, która mówi o awarii jak o liście kontrolnej, a nie jak o hańbie. Na końcu zaś sojusze, bo nikt nie składa parasola na jednokontynentalny deszcz.

Zakaz antropomorfizacji: AI jako narzędzie, nie osoba

Celem nie jest budowa substytutu osoby, lecz konstrukcja narzędzia. Człowiek nie potrzebuje bowiem cyfrowego rywala w samoświadomości, lecz protezy dla własnych kompetencji. Dopiero gdy sztuczna

inteligencja przestaje udawać podmiot, możemy ją rozliczać jako produkt i usługę. To rozróżnienie przywraca klarowne granice prawa, ekonomii i etyki.

Technodyplomacja: globalne sojusze na rzecz bezpieczeństwa

Gdyby kontynenty miały własne, domyślne heurystyki, ich logika byłaby uderzająco odmienna. Europa ufa regułom i procedurom, traktując je jak humanistyczny pancerz chroniący przed chaosem. Ameryka pokłada wiarę w przedsiębiorczości i skali, zakładając, że wielkość sama w sobie jest rozwiązaniem dla poważnych ryzyk. Azja z kolei ufundowała swoją przewagę na państwie strategicznym i długim horyzoncie czasowym, w którym innowacja staje się narzędziem racji stanu. Afryka, najmłodszy demograficznie kontynent, postrzega technologię dwojako: jako obietnicę skrótu rozwojowego i zarazem jako narzędzie cudzej dominacji. A jednak pod powierzchnią tych różnic reguła pozostaje niezmienna. Kto nie potrafi łączyć otwartości z powściągliwością, ten albo tonie w chaosie, albo buduje więzienie na własnym brzegu.

> W epoce, gdy granice między twórcą a narzędziem zacierają się, a państwo staje w obliczu bezprecedensowych wyzwań, kluczowe staje się pytanie o to, czy potrafimy zbudować architekturę odpowiedzialności, zanim technologia zbuduje nas. Czy uda nam się przekształcić nieokiełznaną falę innowacji w źródło powszechnego dobrobytu, czy też zostaniemy zmieceni przez jej potęgę? Przyszłość zależy od naszej zdolności do połączenia otwartości z rozwagą, innowacji z etyką, i technologii ze społeczną sprawiedliwością.

Inspiracje

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

2023 | Crown

Brytyjski przedsiębiorca zajmujący się sztuczną inteligencją; prezes Microsoft AI. Współzałożyciel DeepMind (przejętego przez Google) i Inflection AI. Kierował wdrażaniem sztucznej inteligencji w DeepMind, integrując ją z produktami Google. Orędownik etycznego rozwoju sztucznej inteligencji.

British AI entrepreneur; CEO of Microsoft AI. Co-founded DeepMind (acquired by Google) and Inflection AI. Led applied AI at DeepMind, integrating AI across Google products. Advocate for ethical AI development.

Microsoft

Literatura uzupełniająca

Książki

Literatura pokrewna (Google Scholar):



[1] "Superintelligence: Paths, Dangers, Strategies"

Nick Bostrom

2014 | Oxford University Press | ISBN: 9780199678112

Google Scholar

Artykuły naukowe

Autorzy przywoływani:

[1] "Guidelines for Artificial Intelligence Containment"

Janos Kramar, Roman V. Yampolskiy, Stuart Russell

2016 | AGI

Google Scholar

[2] "Learning the preferences of uncertain humans with inverse decision theory"

Cassidy Laidlaw, Stuart Russell

2021 | arXiv:2106.10394

Google Scholar

[3] "The Impossibility of AI Containment: Logical, Mathematical, and Computational Limits to Control"

Sawsan Haider

2023 | PhilSci-Archive

Google Scholar

APA ONE Publishing System

Fundacja Dobre Państwo © 2026

Article ID: a28072ea-7f23-42d8-91e8-28d7557bd7c2