Statystyka po utracie niewinności: Nowy rygor w dobie AI

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje ewolucję statystyki w obliczu rozwoju sztucznej inteligencji, wskazując na konieczność przejścia od tradycyjnej inferencji do systemów hybrydowych. Autor, nawiązując do prac Tianyu Zhana, omawia rolę głębokich sieci neuronowych jako narzędzi wspierających, a nie zastępujących myślenie statystyczne. Kluczowym pojęciem jest safeguarding – zestaw zabezpieczeń chroniących integralność badań klinicznych przed błędami algorytmicznymi. Tekst podkreśla znaczenie infrastruktury obliczeniowej oraz nowych wytycznych regulacyjnych, takich jak ICH E20, w utrzymaniu rygoru metodologicznego. Statystyka w dobie AI staje się dyscypliną inżynieryjną, gdzie zarządzanie ryzykiem i błędem pierwszego rodzaju wymaga precyzyjnej architektury systemów decyzyjnych i prespecyfikacji ilościowej.

This article examines the evolution of statistics in the face of artificial intelligence, pointing to the need to shift from traditional inference to hybrid systems. Drawing on the work of Tianyu Zhan, the author discusses the role of deep neural networks as tools that support, rather than replace, statistical thinking. A key concept is safeguarding—a set of safeguards protecting the integrity of clinical trials from algorithmic errors. The text emphasizes the importance of computational infrastructure and new regulatory guidelines, such as ICH E20, in maintaining methodological rigor. In the age of AI, statistics is becoming an engineering discipline, where managing risk and type I error requires precise decision-making system architecture and quantitative pre-specification.

Współczesna statystyka przechodzi fundamentalną transformację, w której tradycyjny rygor metodologiczny zderza się z potęgą obliczeniową głębokich sieci neuronowych. Tianyu Zhan w swoim projekcie dowodzi, że sztuczna inteligencja nie musi być rywalem klasycznego wnioskowania, lecz może stać się kluczowym komponentem w budowie

nowoczesnych systemów hybrydowych. Autor odrzuca mit AI jako magicznej wyroczni, proponując wizję statystyka jako architekta inferencji, który wykorzystuje sieci neuronowe do optymalizacji tam, gdzie teoria zawodzi, przy jednoczesnym zachowaniu nienaruszalnych własności granicznych. Kluczem do tej zmiany jest przejście od polegania na dżentelmeńskiej cnocie badacza ku konstytucjonalizacji procedur, w których pre-specyfikacja ilościowa i sztywne ramy obliczeniowe zabezpieczają integralność badania przed pokusami p-hackingu czy uznaniowością. W dobie, gdy infrastruktura obliczeniowa staje się nową konstytucją wiedzy, statystyka traci swoją niewinność, zyskując w zamian nową, cyfrową podmiotowość. Artykuł ten jest wezwaniem do porzucenia technologicznego mesjanizmu na rzecz dojrzałej ekonomii dowodu, w której technologia służy rygorowi, a nie go zastępuje, tworząc fundamenty dla nauki odpornej na nieprzewidywalność rzeczywistości.

SŁOWA KLUCZOWE

statystyka | sztuczna inteligencja | głębokie sieci neuronowe | safeguarding | wnioskowanie statystyczne | badania kliniczne | prespecyfikacja ilościowa | integralność badania | infrastruktura obliczeniowa | estymanda | adaptacyjność | wytyczne ICH E20 | błąd pierwszego rodzaju | estymator zespołowy | rygor metodologiczny

Statystyka w dobie AI: od zegarmistrza do inżyniera mostu

W dziejach nauki zdarzają się momenty, w których zmienia się nie tylko zestaw dostępnych narzędzi, lecz sama definicja roztropności metodologicznej. Współczesna statystyka znajduje się właśnie w takim punkcie, w którym dawne pewniki ustępują miejsca nowej, znacznie bardziej brutalnej rzeczywistości danych. Nie chodzi tutaj o banalny spór między tradycyjną inferencją a uczeniem maszynowym, gdyż ten konflikt w swojej najprostszej postaci stał się już anachroniczny. Prawdziwe pytanie dotyczy sposobu zachowania rygoru wnioskowania w świecie, który stał się zbyt złożony dla eleganckich, tablicowych rozwiązań. Statystyka nie utraciła rozumu, lecz bezpowrotnie utraciła swoją niewinność. Musimy zatem zapytać, jak zarządzać prawdą w epoce, w której czarna skrzynka kusi obietnicą wszechwiedzy, ale nie oferuje żadnych gwarancji.

Projekt Tianyu Zhana jawi się w tym kontekście jako propozycja intelektualnie odświeżająca i metodologicznie niezwykle śmiała. Autor nie proponuje kapitulacji klasycznego warsztatu przed potęgą sztucznej inteligencji, lecz sugeruje jego gruntowne dozbrojenie. Głębokie sieci neuronowe nie mają tutaj zastępować myślenia statystycznego, co byłoby błędem poznawczym najwyższego rzędu. Mają one raczej pracować jako wyspecjalizowane moduły pomocnicze w obszarach, w których klasyczne rozwiązania okazują się analitycznie nieosiągalne lub zbyt kosztowne czasowo. To podejście pozwala na zachowanie kontroli tam, gdzie architektura danych staje się zbyt krucha dla tradycyjnych narzędzi. W ten sposób statystyka zyskuje nową siłę, nie tracąc przy tym swojej krytycznej duszy.

W klasycznej kulturze statystycznej centrum ciężkości spoczywało na jawnie zapisanej strukturze modelu oraz na kontrolowanym rozkładzie błędów. Ceniono elegancję dowodu matematycznego i pełną interpretowalność każdego parametru, co dawało badaczowi poczucie absolutnego panowania nad procesem. W nowym paradygmacie ciężar przesuwa się jednak ku architekturze obliczeniowej oraz ku precyzyjnej konstrukcji danych uczących. Nie oznacza to destrukcji dawnego ładu, lecz raczej jego bolesne urealnienie w obliczu cyfrowego chaosu. Statystyk przestaje być jedynie strażnikiem rozkładu, a staje się architektem złożonego systemu decyzyjnego. Musi on teraz jednocześnie liczyć, przewidywać i chronić integralność badania przed zakusami algorytmicznej arbitralności.

Infrastruktura obliczeniowa, czyli zasoby sprzętowe i moc procesorów traktowane jako fundament współczesnej metodologii, przesuwa punkt ciężkości z samego modelu na ustrój obliczeń. Jeżeli dawny badacz przypominał precyzyjnego zegarmistrza, to nowy musi stać się raczej inżynierem mostu w strefie o wysokiej aktywności sejsmicznej. Nie wystarczy już, że konstrukcja teoretyczna jest piękna i spójna wewnętrznie na poziomie matematycznego abstraktu. Musi ona przede wszystkim wytrzymać drgania rzeczywistości oraz nieprzewidywalne wstrząsy generowane przez napływające strumienie danych. Najciekawszym pojęciem w tym nowym słowniku nie jest wcale samo deep learning, lecz safeguarding. Safeguarding, czyli praktyka zabezpieczania metod statystycznych przed błędami generowanymi przez świat i model, pozwala na utrzymanie rygoru w systemach hybrydowych.

Słowo to posiada posmak prawny, ekonomiczny i antropologiczny jednocześnie, co nadaje mu niemal konstytucyjną rangę w świecie nauki. Wymiar prawny dotyczy gwarancji proceduralnych, które ograniczają uznaniowość badacza i chronią proces przed pokusą manipulacji wynikami. Aspekt ekonomiczny odnosi się do zarządzania ryzykiem błędu oraz kosztami decyzji podejmowanych w warunkach głębokiej niepewności. Wreszcie wymiar antropologiczny odsyła nas do starego pytania o to, kto naprawdę sprawuje władzę nad ostatecznym wynikiem badania. Czy

jest to człowiek, algorytm, czy może ukryty układ ich wzajemnych, często nieuświadomionych współzależności. Dyskusja o metodach wspieranych przez sieci neuronowe staje się zatem rozmową o ustroju poznania.

Propozycja Zhana koncentruje się na wnioskowaniu statystycznym, które w klasycznym języku sprowadza się do testowania hipotez i estymacji punktowej. Na papierze te zadania wyglądają niewinnie, lecz w praktyce adaptacyjnych badań klinicznych potrafią zamienić się w bagno o wielu poziomach lepkości. Gdy projekt badania pozwala na zmianę reguł w trakcie jego trwania, czyste formuły podręcznikowe przestają wystarczać. Randomizacja zaczyna reagować na napływające sygnały skuteczności, co wprowadza do układu niebezpieczną dynamikę i ryzyko obciążenia. W takich warunkach zachowanie nominalnej kontroli błędu pierwszego rodzaju staje się problemem nie tyle rachunkowym, ile wręcz ustrojowym. System musi udowodnić, że każda adaptacja nie jest jedynie przebraniem dla metodologicznej samowoli.

Regulatorzy w Stanach Zjednoczonych i Europie od lat podkreślają, że adaptacyjność jest dopuszczalna, a nawet wysoce pożądana. Musi ona jednak pozostać uprzednio zaplanowana i uzasadniona w ramach rygorystycznej analizy ryzyka operacyjnego. Najnowszym wyrazem tego dążenia jest projekt wytycznych ICH E20, rozwijany wspólnie przez EMA oraz FDA. Dokument ten akcentuje konieczność planowania i interpretacji wyników w rygorze potwierdzającym, co ma zapobiegać interpretacyjnemu chaosowi. Właśnie tutaj pojawia się miejsce na pre-specyfikację ilościową, czyli praktykę polegającą na przeniesieniu kluczowych elementów zmienności do fazy przed badaniem poprzez sztywne reguły. Dzięki temu ograniczamy uznaniowość analityka i budujemy fundament pod rzetelne wnioskowanie.

Propozycja dwuetapowego konstruowania testów za pomocą sieci neuronowych jest czymś znacznie więcej niż tylko zgrabną inżynierską sztuczką. Pierwsza sieć pełni w tym układzie rolę konstruktora statystyki testowej, optymalizując ją pod kątem specyfiki danych. Druga sieć natomiast wyznacza wartości krytyczne w taki sposób, aby bezwzględnie zachować założony poziom błędu. Ujęte w ten sposób uczenie głębokie nie jest rywalem dla tradycyjnego hipotezowania, lecz narzędziem budowy statystyki tam, gdzie teoria nie dostarcza już zamkniętej formy rozwiązania. To bardzo ważna korekta wobec popularnej mitologii AI, która często obiecuje nam magiczne rozwiązania bez żadnych kosztów. Tutaj sieć nie wie lepiej, lecz jedynie sprawniej aproksymuje funkcje wskazane przez człowieka.

Sztuczna inteligencja nie ustanawia tutaj nowej normy inferencji, lecz wykonuje zadanie w obrębie granic wyznaczonych przez statystykę. To rozróżnienie wydaje się na pierwszy rzut oka abstrakcyjne, lecz w rzeczywistości ma charakter fundamentalny dla bezpieczeństwa nauki. Decyduje ono bowiem o tym, czy model pracuje jako lojalny sługa rygoru, czy jako jego

niebezpieczny uzurpator. Estymacja punktowa odsłania drugi wymiar tego sporu, szczególnie widoczny w badaniach z randomizacją uwarunkowaną odpowiedziami. W takich konstrukcjach prosty estymator oparty na naiwnym łączeniu danych z różnych etapów może być poważnie obciążony. To nie jest drobna usterka techniczna, lecz sytuacja, w której intuicja zdradza samą logikę naukowego eksperymentu.

Zhan odpowiada na to wyzwanie, zestawiając ze sobą dwa odmienne tryby działania algorytmicznego. Pierwszy z nich to estymator zespołowy, w którym sieć uczy się wag dla kombinacji spójnych estymatorów bazowych. Drugi polega na estymatorze bezpośrednim, gdzie sieć próbuje samodzielnie aproksymować ostateczną wartość końcową. Z przedstawionego porównania wyłania się dość trzeźwa i nieco konserwatywna prawda o naturze modelowania. Metoda bezpośrednia może czasem osiągać niższy błąd średniokwadratowy, ale bywa jednocześnie bardziej narażona na ryzyko dużego obciążenia poza zakresem treningowym. Metoda zespołowa jest bezpieczniejsza, ponieważ jej struktura zachowuje stały kontakt z klasycznymi fundamentami statystyki i posiada lepsze własności graniczne.

Własności graniczne, czyli statystyczne gwarancje zachowywane nawet przy niedoskonałościach modelu, zapewniają bezpieczeństwo wnioskowania w złożonych systemach hybrydowych. Statystyka, jak rzadko która dziedzina wiedzy, uczy nas głębokiej pokory wobec nadmiaru talentu i technicznej błyskotliwości. To, co w pierwszym odruchu wydaje się bardziej widowiskowe, nie zawsze okazuje się godne naszego zaufania. Jednym z najcenniejszych wkładów tej linii myślenia jest przywrócienie powagi pojęciu integralności badania. Integralność badania, czyli sposób zaprojektowania procedur tak, aby wynik nie był podatny na manipulację po fakcie, gwarantuje obiektywizm procesu. W publicystyce technologicznej słowo to bywa nadużywane do tego stopnia, że przypomina raczej marketingowe perfumy niż twardą zasadę.

Tutaj jednak integralność staje się kategorią ustrojową, a nie tylko moralną dekoracją. To nie jest księgowość, to jest polityczna ekonomia wnioskowania, gdzie cnotę badacza zastępujemy solidną instytucją. Konstytucjonalizacja procedury, czyli budowanie architektury badawczej narzucającej dyscyplinę, ogranicza swobodę improwizacji na danych. Architekt inferencji, czyli nowoczesny statystyk jako projektant problemu i strażnik nienegocjowalnych zasad, definiuje unikalną wartość ludzkiej inteligencji. W świecie zdominowanym przez automatyzację to właśnie on pilnuje, aby algorytmy nie stały się epistemologicznymi oszustami w garniturach z TensorFlow. Nauka bywa bowiem najbezpieczniejsza wtedy, gdy zamiast ufać ślepo w moc obliczeniową, potrafimy ją okiełznać za pomocą rygorystycznie zaprojektowanych struktur.

Mechanizacja uczciwości: AI jako gwarant rygoru badawczego

W badaniach statystycznych i klinicznych integralność badania, czyli praktyka projektowania procedur tak, aby wynik nie był podatny na manipulację, nie jest kwestią dobrych chęci, lecz rygorystyczną architekturą. Chodzi o takie ustawienie reguł adaptacji i procesów analitycznych, aby ostateczny rezultat pozostał odporny na dopasowywanie decyzji do zaobserwowanych już trendów czy miękkie odmiany p-hackingu. Zhan rzuca tu wyzwanie tradycji, wprowadzając kluczowe rozróżnienie między prespecyfikacją jakościową a ilościową. Pierwsza to jedynie obietnica złożona w dokumentacji, swoisty protokół intencji, który badacz spisuje przed przystąpieniem do pracy. Druga idzie znacznie dalej, wymagając, aby algorytm działający na właściwych danych był całkowicie deterministyczny, ponieważ jego parametry zostały wyuczone i zamrożone jeszcze przed napływem danych. To przejście od deklaracji do twardej struktury jest znacznie bardziej radykalne, niż mogłoby się wydawać na pierwszy rzut oka.

Prespecyfikacja jakościowa jest w istocie tylko dżentelmeńską umową, podczas gdy prespecyfikacja ilościowa staje się nienaruszalną architekturą wnioskowania. A nauka, podobnie jak konstytucja, bywa najbezpieczniejsza wtedy, gdy cnotę zastępuje instytucją, budując systemy odporne na ludzkie słabości. W tym miejscu objawia się jedna z najbardziej nieoczywistych korzyści płynących z wykorzystania głębokich sieci neuronowych w rygorystycznej analityce. Choć popkulturowa narracja o sztucznej inteligencji karmi nas obrazami nieprzewidywalnego żywiołu i magicznej emergencji, w statystyce wytrenowany model DNN może pełnić rolę komponentu deterministycznego. Zapisany w plikach parametrów i uruchamiany w ściśle określonym reżimie, staje się on mechanizmem ograniczającym dowolność interpretacyjną. To nie jest jałowa automatyzacja, lecz mechanizacja uczciwości proceduralnej, która chroni proces poznawczy przed chaosem.

Dzięki takiemu podejściu badacz nie musi już ulegać pokusie wielokrotnego uruchamiania losowego bootstrapu czy skomplikowanych łańcuchów Markowa na danych, które swoim kształtem aż proszą się o konkretną interpretację. Otrzymuje on wcześniej ustaloną funkcję, która operuje na nowych informacjach bez pozostawiania pola dla interpretacyjnych machinacji czy intelektualnego samousprawiedliwienia. Oczywiście ten technologiczny ideał nie jest panaceum na wszystkie bolączki współczesnej nauki i nie likwiduje błędów w danych wejściowych. Nie naprawi on źle zaprojektowanego eksperymentu ani nie zwolni zespołu z obowiązku rzetelnej walidacji wyników. Niemniej jednak znacząco zawęża przestrzeń uznaniowości tam, gdzie w codziennej praktyce

badawczej najłatwiej o drobne, niemal niezauważalne oszustwa. Statystyka nie utraciła rozumu, lecz utraciła niewinność, zastępując zaufanie do analityka rygorem zamrożonego kodu.

Współczesny paradygmat regulacyjny, wyznaczany przez wytyczne ICH E9 oraz ich kluczowy dodatek E9(R1), porusza się dokładnie w tym samym kierunku co propozycje Zhana. Nowe standardy uczyniły kwestię precyzyjnego zdefiniowania estymandu, czyli teoretycznej wielkości opisującej efekt leczenia, czymś znacznie bardziej formalnym niż w ubiegłych dekadach. Rzecz nie sprowadza się już wyłącznie do technicznego pytania o to, jak policzyć końcowy wynik w arkuszu kalkulacyjnym. Chodzi o utrzymanie absolutnej spójności między celem badania, wybranym estymatorem a analizami wrażliwości, które testują odporność wniosków na drobne zmiany założeń. Wszak w tym świecie metody wspierane przez DNN mają sens tylko o tyle, o ile pozostają w służbie tego porządku, a nie obok niego. Jeżeli model zwiększa moc statystyczną, ale jednocześnie zaciemnia przedmiot wnioskowania, staje się narzędziem poznawczo podejrzanym.

Jednym z najtrwalszych błędów popełnianych przez teoretyków metodologii jest traktowanie czasu obliczeniowego jako czysto technicznego kosztu, który nie wpływa na samą strukturę wiedzy. To niebezpieczne złudzenie, bowiem czas obliczeniowy posiada status epistemologiczny, ekonomiczny, a wreszcie głęboko etyczny. Epistemologiczny, ponieważ to on wyznacza granice tego, jakie procedury są w ogóle wykonalne i jak głęboko możemy zbadać czułość uzyskanego wyniku. Ekonomiczny, gdyż przelicza się on bezpośrednio na budżet projektu, zużycie infrastruktury oraz opóźnienie kluczowych decyzji biznesowych czy medycznych. Etyczny zaś dlatego, że w badaniach klinicznych każda zwłoka w dostarczeniu wiarygodnej analizy ma realne konsekwencje dla pacjentów czekających na nową terapię. Czas nie jest więc tylko tłem, lecz aktywnym uczestnikiem procesu budowania naukowej pewności.

Zhan bardzo mocno akcentuje potrzebę przeniesienia ciężaru pracy obliczeniowej na etap przygotowawczy, co pozwala na błyskawiczne przeprowadzenie analizy w momencie napływu właściwych danych. Model wytrenowany wcześniej na danych syntetycznych sprawia, że oszczędność czasu przestaje być jedynie efektem ubocznym, a staje się jednym z centralnych uzasadnień nowej metody. Takie rozumowanie doskonale współbrzmi ze współczesnym pejzażem infrastrukturalnym, w którym potężne jednostki GPU i ekosystemy takie jak CUDA współkonstytuują granice naukowego poznania. Tam, gdzie dawna statystyka mogła jeszcze wyobrażać sobie procesor jako neutralne, przezroczyste tło, dzisiejsza metodologia musi przyjąć do wiadomości rolę infrastruktury obliczeniowej. To nie jest wątek poboczny, lecz polityczna ekonomia wnioskowania, w której dostęp do mocy obliczeniowej definiuje granice tego, co uznajemy za możliwe.

Algorytmy jako obywatele infrastruktury i strażnicy dowodu

Gdy badacz mówi o czasie obliczeniowym, nie opisuje jedynie parametrów technicznych, lecz dotyka samej istoty władzy i geopolityki dostępu do narzędzi współczesnego poznania. Artykuł "Dobre Państwo" słusznie zauważa, że jednostki GPU przestały być zwykłymi komponentami, stając się infrastrukturą obliczeniową, czyli fundamentem metodologicznym porównywalnym z siecią energetyczną. To przesunięcie ma fundamentalne znaczenie dla statystyki wspieranej przez głębokie sieci neuronowe, gdyż przenosi ciężar dyskusji z elegancji samego modelu na ustrój obliczeń. Codzienność naukowa w tym kontekście nabiera cech niemal literackiego absurdu, w którym wzniosłe teorie zderzają się z brutalną i prozaiczną materią sprzętową. Wyobraźmy sobie uczonego wygłaszającego z namaszczeniem wykład o pokorze wobec danych, którego projekt zostaje nagle przerwany nie przez błąd logiczny, lecz przez fakt, że las losowy (Random Forest) pożarł całą dostępną pamięć.

Oto współczesna scena poznania, gdzie Homer spotyka administratora klastra, a Arystoteles z rezygnacją wpatruje się w logi błędów systemowych. Statystyk, chcąc zachować resztki zawodowej godności, musi dziś postrzegać algorytm nie jako abstrakcyjną funkcję, lecz jako pełnoprawnego obywatela infrastruktury. Przewaga metod opartych na gradientach nad tradycyjnymi alternatywami nie wynika z przelotnej mody technicznej, lecz z nowej relacji między złożonością zadania a praktyczną wykonalnością procesu. Statystyka nie utraciła rozumu, ale z pewnością utraciła niewinność, stając się integralną częścią większego, politycznego aparatu technologicznego. W tej nowej rzeczywistości las losowy przychodzi z drzewostanem tak okazałym, że administrator serwerowni zaczyna traktować badacza jak cybernetycznego włamywacza. Przewaga ta dotyczy przede wszystkim relacji między teoretyczną głębią a możliwością jej faktycznego przetestowania w skończonym czasie.

Nic dziwnego, że adaptacyjne badania kliniczne, czyli projekty pozwalające na modyfikację parametrów na podstawie danych cząstkowych, stają się poligonem dla tej nowej racjonalności. Adaptacyjność jest tu zarazem wielką szansą na etyczniejszy przebieg testów, jak i niebezpieczną pokusą do oportunistycznego manipulowania wynikami pod presją czasu. Szansa polega na możliwości skrócenia ścieżek decyzyjnych i lepszej alokacji pacjentów, co bezpośrednio przekłada się na bezpieczeństwo uczestników i efektywność kosztową. Pokusa natomiast kryje się w integralności badania, czyli sposobie projektowania procedur tak, aby wynik nie był podatny na manipulację po fakcie. Wytyczne FDA z 2019 roku oraz projekt ICH E20 dążą do skodyfikowania zasad, które zamienią analityczną anarchię w uporządkowany teren normatywny. Nie jest to więc

obszar radosnej improwizacji, lecz przeciwnie, przestrzeń o rosnącym stopniu rygorystycznego, instytucjonalnego uporządkowania.

W tym miejscu podejście Zhana ujawnia swoją praktyczną siłę, oferując narzędzia do rygorystycznego zarządzania niepewnością w dynamicznym środowisku badawczym. Badania z ponownym szacowaniem wielkości próby są zbyt skomplikowane dla klasycznych wzorów, a jednocześnie zbyt istotne, by pozwolić sobie na analityczną improwizację. Jeśli model potrafi wyznaczyć adekwatne wartości krytyczne przy zachowaniu własności brzegowych, czyli statystycznych gwarancji rygoru nawet przy niedoskonałościach aproksymacji, zyskujemy dojrzałą ekonomię dowodu. Architekt inferencji, czyli nowoczesny statystyk projektujący ramy problemu i strażnik nienegocjowalnych zasad wnioskowania, staje się wówczas kluczowym ogniwem procesu. Taka obietnica jest wyjątkowo atrakcyjna z perspektywy ekonomii zdrowia i polityki lekowej, gdzie każda przyspieszona decyzja może ratować ludzkie życie. Zyskujemy coś znacznie cenniejszego niż sama szybkość obliczeń; otrzymujemy metodologię, która potrafi udźwignąć ciężar odpowiedzialności za błąd.

Należy jednak zachować zdrowy sceptycyzm wobec twierdzeń o uniwersalnej wyższości jakiejkolwiek metody, pamiętając o systemowych lukach w samych danych. Współczesne badanie kliniczne to nie tylko maszyna do generowania p-wartości, lecz instrument poznania ciał zróżnicowanych biologicznie i społecznie. Jak pokazuje publikacja o luce w danych, historyczne oparcie medycyny na niejawnie męskim modelu normy nie jest drobnym błędem tabelarycznym, lecz systemowym zafałszowaniem racjonalności. To nie jest zwykła księgowość, to jest polityczna ekonomia wnioskowania, w której błąd statystyczny nie rozkłada się równo między grupami społecznymi. Jeżeli model uczy się na tak skrzywionej i źle zbudowanej rzeczywistości, jego obliczeniowa elegancja może jedynie utrwalić pierwotną asymetrię. Epistemologiczny oszust w garniturze z TensorFlow może wówczas bardzo szybko poprowadzić nas ku wyrafinowanym, lecz głęboko błędnym wnioskom.

Dlatego zabezpieczanie (safeguarding), czyli budowanie systemowych zabezpieczeń wokół modelu, nie może ograniczać się jedynie do ochrony poziomu błędu pierwszego rodzaju. Musi ono obejmować szeroko pojętą konstytucjonalizację procedury, która narzuca dyscyplinę i ogranicza swobodę improwizacji na surowych danych poprzez sztywne reguły. Pytanie nie brzmi, czy zaawansowane modele bywają skuteczne, lecz czy potrafimy obudować je takimi ograniczeniami, by nie zniszczyły wartości poznawczej badania. Problem danych spoza zakresu nie jest techniczną anomalią, lecz naturalnym prawem empirii, która uwielbia pokazywać badaczom, że ich zbiory treningowe były zbyt grzeczne. Rozszerzanie zakresów danych uczących to tylko pierwsza linia obrony przed błędem, podczas gdy linia druga musi być znacznie głębsza i bardziej systemowa.

Nauka, podobnie jak konstytucja, bywa najbezpieczniejsza wtedy, gdy cnotę badacza zastępuje solidną i nienaruszalną instytucją metody. Bez tego zabezpieczanie pozostanie jedynie pustym hasłem, a my ryzykujemy budowę bardzo szybkich autostrad prowadzących wprost do metodologicznej przepaści.

Statystyka w maszynowni: Hybrydowy rygor obliczeniowy

Fundamentem nowoczesnego podejścia do modelowania jest budowa ram statystycznych, które zachowują swoje właściwości graniczne, czyli matematyczne gwarancje rzetelności, nawet w sytuacjach kryzysowych. Dzieje się tak zwłaszcza wtedy, gdy model nie potrafi wyznaczyć idealnych wag lub gdy jego aproksymacja drastycznie traci na precyzji. To jest właśnie ten kluczowy moment, w którym ludzka inteligencja okazuje się nie ozdobnym dodatkiem do algorytmu, lecz suwerenem całego układu poznawczego. Tylko ekspert posiada bowiem kompetencje, by zdecydować, jakie ograniczenia muszą zostać wpisane w samą konstrukcję estymatora, aby nie dryfował on w stronę czystego szumu. On również wybiera statystyki bazowe, projektuje wejścia w oparciu o wiedzę dziedzinową i organizuje proces walidacji tak, by model nie zamienił się w epistemologicznego oszusta w garniturze z TensorFlow. Projekt Tianyu Zhana jest pod tym względem ciekawie konserwatywny, gdyż nie ulega modnemu mitowi o pełnej autonomii sztucznej inteligencji. Przeciwnie, autor konsekwentnie oddaje nadrzędną rolę ludzkiemu rozeznaniu, co stanowi trzeźwą odtrutkę na współczesne bolączki metodologiczne.

Współczesny dyskurs o sztucznej inteligencji cierpi bowiem na chorobę infantylnego determinizmu technologicznego, który bezkrytycznie zakłada, że proces optymalizacji jest tożsamy z posiadaniem wiedzy. Nic bardziej mylnego, gdyż model optymalizuje wyłącznie to, co zostało mu narzucone w funkcji celu oraz to, na co pozwalają mu dostarczone dane. Maszyna nie zna kategorii dowodu, nie rozumie sensu odpowiedzialności regulacyjnej i nie posiada moralnej intuicji dotyczącej konsekwencji błędów. Niewielkie przesunięcie błędu w jedną stronę może mieć przecież ogromne znaczenie dla pacjentów, inwestorów oraz całego porządku zaufania publicznego. Stąd bierze się absolutna konieczność zachowania hierarchii, w której najpierw następuje badanie, potem określenie estymandu, a dopiero na samym końcu dobór algorytmu. Ta kolejność nie jest jedynie konserwatywnym kaprysem starej szkoły, lecz warunkiem, by sztuczna inteligencja nie stała się po prostu szybszą drogą do bardziej elegancko opakowanej nierzetelności. To nie jest księgowość, to jest ekonomia polityczna wnioskowania, w której stawką jest wiarygodność nauki.

Skoro omówiliśmy już ustrój wiarygodności, musimy teraz zejść poziom niżej, bezpośrednio do dusznej maszynowni obliczeń numerycznych. Tam bowiem rozstrzyga się znacznie więcej kwestii merytorycznych, niż skłonni są przyznać humaniści statystyki, często bujający w obłokach czystej teorii. W świecie klasycznej dydaktyki wciąż podtrzymuje się wygodne złudzenie, że numeryka jest jedynie pokorną usługą wobec wielkiej myśli matematycznej. Według tego schematu najpierw powstaje genialne rozumowanie, a dopiero potem rachunek je realizuje, nie wnosząc nic nowego do samej istoty problemu. W praktyce współczesnych badań klinicznych oraz analiz wysokowymiarowych ten tradycyjny porządek dawno już uległ całkowitemu odwróceniu. To, co da się policzyć w sensownym czasie przy akceptowalnym zużyciu pamięci, zaczyna współdecydować o tym, co w ogóle uznajemy za metodologicznie poważne. Statystyka nie utraciła rozumu, ale z całą pewnością utraciła niewinność, stając się zakładnikiem możliwości sprzętowych.

Musimy dziś pogodzić się z faktem, że rygor naukowy nie mieszka wyłącznie w eleganckim dowodzie, lecz także w konkretnej architekturze obliczeniowej. Teza Zhana jest pod tym względem bezlitosna i zarazem wyzwalająca, ponieważ głębokie uczenie nie ma tutaj jedynie ozdabiać starych metod numerycznych. Ma ono raczej reorganizować relację między trudną funkcją, jej aproksymacją, gradientem oraz kosztami przeszukiwania ogromnej przestrzeni potencjalnych rozwiązań. Książka obejmuje nie tylko inferencję i estymację, lecz także optymalizację, interpretowalność i konstrukcję procedur chroniących integralność badania przed manipulacjami. Sugestywny jest tu zwłaszcza wątek optymalizacji bez informacji o gradiencie, który dotyczy wielu zadań praktycznych w projektowaniu złożonych eksperymentów. Prowadzą one często do funkcji celu, które są niewygodne, nieregularne, niewypukłe i, co tu dużo mówić, wyjątkowo niewdzięczne poznawczo dla analityka.

Tradycyjne algorytmy bezgradientowe potrafią się w takim trudnym terenie poruszać, ale czynią to zazwyczaj powoli, niestabilnie i z ogromnym zużyciem zasobów. Pomysł, by taką oporną funkcję celu zastąpić dobrze wytrenowaną siecią neuronową, nie jest technologiczną fanaberią, lecz próbą wprowadzenia do tej dżungli struktury różniczkowalnej. Wytrenowana sieć DNN staje się sprawnym aproksymatorem, który udostępnia gradienty, a przez to pozwala uruchomić znacznie szybsze algorytmy oparte na pochodnych. Następnie uzyskane rozwiązanie może zostać jeszcze dopracowane metodą bezgradientową, aby skutecznie zredukować ryzyko błędu aproksymacji. To podejście nie ma nic wspólnego z metafizyką AI, gdyż stanowi wyraz bardzo trzeźwej ekonomii liczenia i szacunku do czasu. Model nie jest tutaj wyrocznią, lecz mostem między nieprzyjazną funkcją a operacyjną wykonalnością całej procedury badawczej.

Właśnie w tym sensie głębokie uczenie wspiera metody numeryczne, zamiast je kolonizować czy zastępować ich fundamenty. Warto zatrzymać się nad logiczną subtelnością tego ruchu, który

przypomina współpracę doświadczonego przewodnika z nowoczesną technologią. Klasyczna optymalizacja bezgradientowa bywa szlachetna jak stary podróżnik, który zna góry, ale porusza się powoli z powodu ciężkiego plecaka i bolących kolan. Sieć neuronowa w tej metaforze nie jest nowym królem gór, lecz kimś, kto sporządza użyteczną mapę terenu, wystarczająco dokładną, by podróżnik przestał błądzić we mgle. Tego rodzaju współpraca między przybliżeniem a rygorem to jedna z najważniejszych lekcji płynących z całej propozycji Zhana. Nauka nie musi wybierać między pięknem teorii a brutalnością inżynierii, lecz może budować układy hybrydowe, w których każde narzędzie wykonuje to, co potrafi najlepiej.

Problem zaczyna się dopiero wtedy, gdy narzędzie pomocnicze zaczyna być czczone jak bóstwo, z czym współczesny dyskurs technologiczny ma spory kłopot. Kłopot ten ma zresztą wymiar czysto ekonomiczny, gdyż czas obliczeniowy nie jest tylko kosztem technicznym, lecz przede wszystkim kosztem alternatywnym. Przelicza się on na opóźnienie kluczowych decyzji, na długość iteracji projektowych oraz na praktyczny zakres walidacji modelu. Jeśli procedura klasyczna potrzebuje dziesiątek godzin pracy, to w realnej organizacji badawczej będzie używana rzadziej i na mniejszej liczbie scenariuszy. Jeżeli zaś model wytrenowany wcześniej pozwala później wykonywać tysiące obliczeń niemal natychmiast, wówczas zmienia się jakość samego procesu poznawczego. Można sprawdzić więcej, szybciej wykryć nieciągłości i wcześniej zidentyfikować ryzyka, które inaczej pozostałyby ukryte pod warstwą obliczeniowej niemocy.

Dlatego Zhan tak mocno akcentuje pre-specyfikację ilościową, czyli praktykę polegającą na przesunięciu kluczowych elementów zmienności do fazy przed badaniem poprzez sztywne reguły. To nie jest zwykła księgowość, lecz ekonomia polityczna wnioskowania, która bardzo dobrze współgra z diagnozami dotyczącymi infrastruktury obliczeniowej. Nowoczesne uczenie głębokie nie jest samą matematyką, lecz splotem architektury sprzętowej, energii oraz globalnej geopolityki mocy obliczeniowej. Dominacja jednostek GPU nie jest estetycznym zwycięstwem pewnej szkoły programowania, lecz reorganizacją warunków tego, co w ogóle da się zrobić szybko i na odpowiedniej skali. Kiedy więc statystyk mówi dziś o symulacji czy walidacji, mówi również nieuchronnie o materialnym tle obliczeń, gdzie Homer spotyka administratora klastra. Problem czasu obliczeniowego ma więc wymiar ustrojowy, gdyż nie każdy zespół dysponuje tą samą infrastrukturą i dostępem do akceleratorów.

W konsekwencji pytanie o metody wspierane przez sieci neuronowe nie jest wyłącznie pytaniem o statystykę, lecz o dostęp do technicznej suwerenności. Każda procedura statystyczna używana do wspierania decyzji klinicznej lub regulacyjnej jest w istocie częścią ładu normatywnego. Integralność badania, czyli sposób zaprojektowania procedur tak, aby wynik nie był podatny na manipulację po fakcie, staje się tu kluczowa. Architekt inferencji, czyli nowoczesny statystyk będący

strażnikiem nienegocjowalnych zasad wnioskowania, musi dziś panować nad tym technologicznym splotem. Konstytucjonalizacja procedury, polegająca na budowaniu architektury badawczej narzucającej dyscyplinę, zastępuje zaufanie do cnoty badacza systemowymi zabezpieczeniami rzetelności. A nauka, podobnie jak konstytucja, bywa najbezpieczniejsza wtedy, gdy cnotę zastępuje instytucją, nawet jeśli tą instytucją jest algorytm zamknięty w rygorystycznych ramach ludzkiego osądu.

Architektura zaufania: Jak ucywilizować algorytmy w dobie AI

W świecie regulowanych algorytmów samo stwierdzenie, że coś „po prostu działa", jest zaledwie uprzejmym wstępem do właściwej rozmowy, a nie jej satysfakcjonującą konkluzją. System musi bowiem funkcjonować w sposób, który daje się racjonalnie uzasadnić, odtworzyć w dowolnym momencie i – co najważniejsze – obronić przed zarzutem czystej arbitralności. Prespecyfikacja ilościowa, czyli praktyka polegająca na sztywnym ustaleniu reguł analizy i parametrów modelu jeszcze przed pierwszym kontaktem z danymi, odpowiada na ten wymóg w sposób wyjątkowo elegancki. Przenosi ona punkt ciężkości z fazy post hoc, gdzie analityk bywa kuszony przez p-hacking i inne formy statystycznego kuglarstwa, do fazy czystego projektu. Można to nazwać swoistą konstytucjonalizacją procedury, gdzie zamiast liczyć na nieskazitelną cnotę badacza, budujemy architekturę ograniczającą jego późniejszą uznaniowość. Statystyka nie utraciła tu rozumu. Utraciła niewinność, zastępując ją rygorem zapisanym w protokole.

W naukach prawnych takie podejście brzmi znajomo, przypominając fundamentalną różnicę między deklaracją dobrych intencji a instytucjonalnym podziałem kompetencji. Zhan nie posługuje się wprawdzie językiem konstytucjonalisty, lecz logika jego wywodu jest uderzająco podobna. Nie chodzi o to, aby badacz był aniołem wolnym od uprzedzeń, bo to postulat z natury nierealny. Wystarczy, że system nie będzie oparty na naiwnym założeniu, że nim jest. Integralność badania, czyli sposób zaprojektowania procedur tak, aby wynik nie był podatny na manipulację po fakcie, staje się tu nienegocjowalnym fundamentem. To nie jest zwykła technika; to ustrój poznania, w którym każda decyzja analityczna musi mieć swoje umocowanie w precyzyjnie zdefiniowanej strukturze, zanim jeszcze na horyzoncie pojawią się pierwsze rekordy z bazy danych.

Interpretowalność bywa dziś w debacie publicznej traktowana jak talizman, magiczne słowo mające egzorcyzmować lęki przed „czarną skrzynką" sztucznej inteligencji. Często ulegamy uproszczonemu podziałowi na modele z natury wyjaśnialne i te beznadziejnie mroczne, co jest dychotomią godną folderu reklamowego, a nie poważnej metodologii. Podejście Zhana do modeli regresji proponuje

rozwiązanie znacznie dojrzalsze, odrzucając pokusę powrotu do modeli ascetycznych tylko po to, by człowiek poczuł się psychicznie komfortowo. Autor sugeruje budowę wielowarstwowych funkcji parametrycznych, których komponenty bazowe wynikają bezpośrednio z wiedzy dziedzinowej. Ich złożoność nie jest pozostawiona samej sobie, lecz podlega surowej ocenie przy użyciu zmodyfikowanej statystyki Mallowsa, która bezlitośnie karze za nadmiarowe parametry nieprzynoszące realnego zysku poznawczego.

Jest to zatem interpretowalność inżynieryjna, a nie dewocyjna. Nie oczekujemy, że model będzie prosty jak przepis na rosół, bo świat, który próbujemy opisać, rzadko bywa tak klarowny. Chodzi o to, by jego elementy były zakotwiczone w głębokim rozumieniu problemu, a poziom komplikacji pozostawał pod ścisłą kontrolą matematyczną. Dla odbiorców regulacyjnych i biznesowych taka logika ma znaczenie fundamentalne. Menedżer, regulator czy sponsor nie potrzebują modelu infantylnie prostego, który gubi niuanse rzeczywistości. Potrzebują narzędzia, którego użycie można obronić przed komisją rewizyjną bez uciekania się do metodologicznej mistyki czy zasłaniania się „intuicją algorytmu". To złudzenie, że prostota jest tożsama z prawdą, bywa w nauce wyjątkowo kosztowne.

Współczesna recepcja tego zagadnienia wyraźnie przesuwa się w stronę skalibrowanej przejrzystości, adekwatnej do ryzyka podejmowanej decyzji i kontekstu dokumentacyjnego. W badaniach klinicznych nie oczekujemy, że każda część matematycznego pipeline'u będzie intuicyjna dla laika, ale każda musi dać się opisać, zwalidować i osadzić w procedurze chroniącej interpretowalność wyniku końcowego. To ważna różnica: wynik badania nie jest prezentem od modelu, lecz produktem systemu, w którym model jest tylko jednym z komponentów. Jeśli komponent cząstkowy jest złożony, można to zaakceptować pod warunkiem, że architektura całości zachowuje możliwość audytu. Tak rozumiana interpretowalność jest znacznie bliższa rzetelnej rachunkowości odpowiedzialności niż szkolnej tęsknocie za prostymi równaniami liniowymi.

Rozdział poświęcony konfrontacji sieci neuronowych z algorytmami takimi jak SVM, Random Forest czy XGBoost stanowi swoisty test uczciwości całego przedsięwzięcia. Gdyby autor ograniczył się do bezkrytycznej apologetyki głębokiego uczenia, moglibyśmy podejrzewać go o zwykły technologiczny mesjanizm. Tymczasem książka traktuje wybór algorytmu jako pewnego rodzaju hiperparametr problemu, co jest podejściem metodologicznie zdrowym. Nie każda procedura wymaga sieci głębokiej, tak jak nie każdy remont wymaga użycia młota pneumatycznego. Pytanie brzmi jednak, jak poszczególne metody zachowują się tam, gdzie stawką jest nie tylko predykcja, lecz również kontrola błędu pierwszego rodzaju, stabilność estymacji i zgodność z wymogami integralności. W tym punkcie obraz zaczyna faworyzować DNN, które w specyficznych reżimach regulacyjnych okazują się paradoksalnie bardziej przewidywalne.

Wyobraźmy sobie, dla rozjaśnienia tej technicznej gęstwiny, salę konferencyjną, w której cztery algorytmy ubiegają się o etat w krytycznym badaniu klinicznym. Random Forest przychodzi z lasem drzew tak okazałym, że ochroniarz natychmiast pyta o pozwolenie budowlane i plan zagospodarowania przestrzennego. SVM rozkłada płaszczyzny separacji z miną matematycznego arystokraty, ale po kilku godzinach medytacji nad dziesięcioma milionami próbek nadal nie wie, gdzie postawić pierwszy krok. XGBoost wbiega żwawo, wszystko wzmacnia, wszystko poprawia, po czym ktoś z końca sali nieśmiało pyta, czy przypadkiem nie napompował trochę za bardzo błędu pierwszego rodzaju w pogoni za wynikiem. DNN natomiast nie wygląda na najbardziej elegancki obiekt w pomieszczeniu, lecz siada do pracy, korzysta z mini-batchy, liczy gradienty i po pewnym czasie mówi, że nie jest cudem, ale przynajmniej nadaje się do dalszej, merytorycznej rozmowy.

Humor tej sceny bierze się stąd, że każdy z tych modeli jest potężny w pewnym rejestrze, lecz nie każdy jest zbudowany do dźwigania tej samej odpowiedzialności procesowej. Z przedstawionych wyników wynika, że alternatywne algorytmy, choć użyteczne, częściej cierpią na problemy kalibracyjne lub ograniczenia zbieżności przy bardzo dużych zbiorach danych. XGBoost, mimo swojej ogromnej mocy obliczeniowej, nie zawsze utrzymuje nominalny poziom błędu typu pierwszego przy konstrukcji statystyk testowych, co w świecie regulowanym jest grzechem ciężkim. SVM w niektórych zadaniach estymacyjnych okazuje się po prostu niewykonalny obliczeniowo, a Random Forest potrafi stać się ofiarą własnej pamięciożerności. To nie są tylko techniczne detale; to bariery, które decydują o tym, czy model może stać się częścią dowodu naukowego.

W publicznej debacie o AI często miesza się dwa fundamentalnie różne pytania. Pierwsze brzmi: który model ma najwyższą trafność predykcyjną w pewnym zamkniętym benchmarku? Drugie natomiast: który model nadaje się do osadzenia w procedurze inferencyjnej obciążonej wymaganiami prawnymi i etycznymi? To nie są pytania tożsame. Model może być znakomity w klasyfikacji marketingowej, a jednocześnie kompletnie zawodzić jako komponent w procedurze, która ma pilnować nominalnego poziomu błędu w adaptacyjnym badaniu klinicznym. Z tego powodu porównania Zhana są tak cenne – uświadamiają nam, że przydatność modelu mierzy się całym pakietem własności: kalibracją, odpornością na zmianę scenariusza i relacją między obciążeniem a błędem średniokwadratowym. To jest naukowa dojrzałość, której często brak w fetyszystycznym zachwycie nad tabelami wyników.

Najmocniejszym momentem analizy poświęconej bezpieczeństwu jest kwestia tego, co dzieje się, gdy rzeczywiste dane wychodzą poza zakres, na którym model był trenowany. W praktyce statystycznej jest to zjawisko tak banalne, że aż skandalicznie często bagatelizowane przez zespoły projektowe, które zachowują się, jakby świat miał obowiązek respektować granice ich symulacji. Tymczasem rzeczywistość empiryczna ma zwyczaj przychodzić nie tam, gdzie ją zapraszano,

niszcząc misternie tkane założenia o stacjonarności procesów. Zhan pokazuje, że w takich warunkach estymator bezpośredni ulega silniejszemu wzrostowi obciążenia niż estymator zespołowy. To przypomnienie starej prawdy: model bardziej elastyczny nie zawsze jest bezpieczniejszy. Elastyczność bez ograniczeń to często po prostu ładniej ubrana forma kruchości, która pęka przy pierwszym kontakcie z nietypową obserwacją.

Własności graniczne, czyli statystyczne gwarancje zachowywane nawet przy niedoskonałościach modelu, stają się tu ostatnią linią obrony. Rozszerzanie zakresów parametrów danych treningowych jest pierwszym ruchem obronnym, ale rzadko bywa wystarczające. Jeśli mechanizm generowania danych użyty do symulacji rozmija się z rzeczywistą strukturą świata, sam szeroki zakres nie uratuje sytuacji przed katastrofą. Potrzebna jest architektura, która zachowa minimalne własności bezpieczeństwa nawet pod błędem modelu. A nauka, podobnie jak konstytucja, bywa najbezpieczniejsza wtedy, gdy cnotę zastępuje instytucją. W tym miejscu rola człowieka jako architekta inferencji zyskuje swój prawdziwy ciężar. Ekspert nie ma być kapłanem podpisującym raport z pracy algorytmu, lecz projektantem granic i strażnikiem nienegocjowalnych zasad wnioskowania.

Jeśli badacz wie z teorii problemu, że pewne statystyki dostateczne redukują złożoność wejścia bez utraty istotnej informacji, powinien je bezwzględnie zastosować. Jeśli rozumie, że dana procedura klasyczna zachowuje fundamentalne własności brzegowe, powinien zbudować wokół niej hybrydę zamiast ją bezmyślnie porzucać na rzecz nowszych trendów. AI nie zrobi tego sama, nie z lenistwa, lecz z braku pojęcia o tym, co jest epistemicznie cenne. Model nie zna pojęcia błędu cywilizacyjnego, człowiek zaś powinien je znać doskonale. Safeguarding jest czymś znacznie szerszym niż technika; to filozofia odpowiedzialności metodologicznej, która zakłada, że każda procedura silna predykcyjnie jest niewinna tylko do chwili spotkania z nieprzewidzianym światem. Dopiero potem okazuje się, czy była narzędziem postępu, czy wyrafinowaną pułapką.

Powróćmy na koniec do badań klinicznych, ponieważ tam cała ta architektura pokazuje swoją najbardziej konkretną, ludzką stawkę. Adaptacyjny projekt badania kusi dziś nie tylko statystyka, ale i ekonomistę, prawnika oraz antropologa instytucji. Ekonomistę, ponieważ obiecuje lepszą alokację zasobów i redukcję kosztu błędnie kontynuowanych projektów. Prawnika, bo wymusza większą formalizację zasad modyfikacji badania i ochronę przed operacyjną stronniczością. Antropologa zaś dlatego, że ujawnia, w jaki sposób nowoczesna medycyna negocjuje zaufanie między pacjentem a algorytmem. W tym polu nie ma nic neutralnego. Nawet sama obietnica, że w drugim etapie większy odsetek pacjentów trafi do obiecującej terapii, jest jednocześnie obietnicą moralną i techniczną, której nie wolno rzucać na wiatr.

Aby tę obietnicę zrealizować uczciwie, trzeba mieć procedurę, która nie sprzedaje etycznego patosu za cenę inferencyjnego chaosu. Projekty takie jak response adaptive randomization są fascynujące, ale i niebezpieczne, ponieważ komplikują estymację i wprowadzają zależności, które mogą produkować obciążenia tam, gdzie naiwny obserwator widzi wyłącznie moralny postęp. To miejsce, w którym statystyka wykonuje jedną ze swoich najbardziej cywilizacyjnych prac. Przypomina nam, że dobra intencja nie zwalnia z obowiązku poprawnego wnioskowania. To nie jest zwykła księgowość danych; to jest polityczna ekonomia wnioskowania, w której stawką jest wiarygodność całej nauki. Ujęcie Zhana zasługuje na uwagę właśnie dlatego, że nie głosi mitu o samowystarczalności sieci neuronowych, lecz stale przypomina o konieczności nadzoru i systemowych zabezpieczeń.

Statystyczny republikanizm: Nowy rygor w dobie AI

Model DNN, czyli głęboka sieć neuronowa służąca do aproksymacji złożonych funkcji, może pełnić rolę estymatora zespołowego, o ile zostanie osadzony w rygorystycznej procedurze bezpieczeństwa. Projekt Zhana nie jest ani technokratycznym cynizmem, ani sentymentalną próbą ratowania etyki, lecz stanowi surową próbę pogodzenia efektywności z odpowiedzialnością. To podejście redefiniuje racjonalność badawczą, odchodząc od ślepej wiary w algorytm na rzecz precyzyjnie zaprojektowanej architektury wnioskowania. Współczesny paradygmat regulacyjny, reprezentowany przez FDA, podąża zresztą w bardzo podobnym kierunku, kładąc nacisk na kontrolę błędów i pełną interpretowalność wyników. W tym sensie statystyka nie utraciła rozumu, lecz utraciła niewinność, zmuszając nas do budowania systemów, które są bezpieczne z definicji, a nie tylko z intencji.

Najnowsze wytyczne FDA dotyczące badań adaptacyjnych wyraźnie wskazują, że innowacja nie musi oznaczać rezygnacji z naukowego rygoru. Projekt ICH E20, ogłoszony równolegle przez EMA i FDA w 2025 roku, potwierdza wysoką wartość projektów elastycznych, o ile służą one rzetelnej ocenie skuteczności i analizie ryzyka. To kluczowy moment, który ostatecznie rozwiewa tani stereotyp o regulatorze jako strażniku skansenu, pragnącym jedynie konserwować przeszłość za wszelką cenę. Instytucje te akceptują nowoczesność, pod warunkiem że zostanie ona przełożona na język wiarygodności, dokumentowalności oraz pełnej przewidywalności procedury. Statystyka wspomagana uczeniem głębokim (deep learning assisted statistics) nie stoi zatem poza prawem, lecz wchodzi w jego ramy, spełniając warunki znacznie surowsze niż zwykła skuteczność predykcyjna.

Nie wolno nam jednak dopuścić, by dyskusja o integralności metody pozostała ślepa na polityczny wymiar gromadzenia i selekcji danych. Tekst „Dobre Państwo", analizujący lukę w danych dotyczących kobiet, stanowi tutaj niezwykle istotny korektyw dla czysto technicznego myślenia o statystyce. Pokazuje on dobitnie, że błąd systemowy nie zawsze objawia się w postaci źle ustawionego poziomu alfa, lecz bywa zakodowany głębiej. Często tkwi on w samym materiale empirycznym, w tym, kogo reprezentują dane historyczne i jakie ciała przez dekady traktowano jako uniwersalną normę. Jeśli metody wspierane przez DNN mają być narzędziem nowoczesnego rygoru, muszą najpierw przyjąć na siebie ciężar tej fundamentalnej krytyki.

Model uczony na stronniczych danych nie staje się neutralny tylko dlatego, że jego matematyczna walidacja wygląda elegancko i profesjonalnie. Może on wręcz skuteczniej utrwalać wcześniejsze asymetrie społeczne, jeśli architektura badania nie przewiduje kontroli reprezentatywności oraz heterogeniczności efektu. Artykuł „Dobre Państwo" trafnie uderza w mit neutralności, wykazując, że dane rzekomo pozbawione płci w praktyce okazują się danymi o mężczyznach. To ostrzeżenie nie jest jedynie poboczną uwagą dla statystyki klinicznej, lecz stanowi jej elementarny test moralny i poznawczy. Bez rozpoznania tych kwestii nowoczesna metodologia staje się jedynie szybszym sposobem reprodukcji naszych wąskich i często błędnych wyobrażeń o normalności.

W tym miejscu antropologia spotyka biostatystykę, ponieważ każdy system ekspercki produkuje określony obraz człowieka, nawet jeśli robi to nieświadomie. Badanie kliniczne nie jest przecież tylko suchą procedurą weryfikacji cząsteczki chemicznej, lecz sposobem ustalenia, kto liczy się jako wiarygodny nosiciel danych. To proces definiowania, jakie różnice między ludźmi uznajemy za istotne i jak szeroko rozumiemy populację, która ma korzystać z przyszłych leków. Dlatego nowoczesny safeguarding, czyli system zabezpieczeń chroniący proces badawczy przed błędami, powinien obejmować odporność na fałszywy uniwersalizm populacyjny. Takie rozszerzenie perspektywy nie osłabia projektu Zhana, lecz czyni go bardziej aktualnym i znacznie bardziej godnym intelektualnej obrony.

Współczesna recepcja tych metod oscyluje między dwiema skrajnymi i równie niebezpiecznymi pokusami, które paraliżują rzetelną dyskusję o przyszłości nauki. Pierwszą jest bezkrytyczny entuzjazm technologiczny, zakładający, że wystarczy uruchomić odpowiednio duży model, by z chaosu wyłoniła się nowa racjonalność. Drugą stanowi konserwatyzm metodologiczny, widzący w każdej architekturze głębokiej niemal definicyjne zagrożenie dla rzetelności i interpretowalności wyników. Projekt Zhana proponuje trzecią drogę, którą można by nazwać statystycznym republikanizmem, czyli ustrojem, w którym władza nad wnioskowaniem jest rozproszona. Nie oddaje on pełnej kontroli ani intuicji badacza, ani autonomii algorytmu, lecz precyzyjnie rozkłada kompetencje między człowieka a maszynę.

W tym układzie klasyczna wiedza statystyczna określa, co ma być chronione, podczas gdy symulacja buduje bezpieczną przestrzeń uczącą dla modelu. DNN aproksymuje to, czego nie umiemy wygodnie wyprowadzić analitycznie, ale mechanizmy walidacji pilnują, by system nigdy nie utracił swoich podstawowych cnót. Prespecyfikacja ilościowa, czyli praktyka przenoszenia kluczowych decyzji analitycznych do fazy przed badaniem, zamraża ścieżkę obliczeń i skutecznie ogranicza uznaniowość. Cały ten system nie jest ani naiwnie klasyczny, ani naiwnie algorytmiczny, lecz rozumie, że złożoność wymaga podziału funkcji. Chodzi o to, by system był bezpieczny nie dlatego, że wierzymy w jego cnotę, lecz dlatego, że jego architektura ogranicza zakres katastrofy.

Podejście to dobrze koresponduje z ruchem trustworthy AI, gdzie audytowalność jako infrastruktura (auditability as infrastructure) staje się fundamentem nowoczesnego zarządzania danymi i procesami. Nie musimy przyjmować wszystkich modnych etykiet, by dostrzec wspólny rdzeń tych starań o wiarygodność i powtarzalność wyników. Kluczowa jest konsekwentna obrona właściwości brzegowych, czyli statystycznych gwarancji zachowywanych nawet przy niedoskonałościach modelu lub utracie precyzji. Badanie może nie uzyskać idealnie najmniejszego błędu, ale nie wolno mu stracić kontroli nad tym, co stanowi minimalny warunek jego wiarygodności. To bardzo głęboka lekcja, przypominająca nam, że w nauce, podobnie jak w polityce, wolność działania wymaga solidnej konstytucji.

Między mocą obliczeniową a prawomocnością inferencyjną

Warto zacząć od demitologizacji popularnego pytania o to, czy sztuczna inteligencja ostatecznie odeśle statystyków na przedwczesną emeryturę. Otóż nie zastąpi ich, o ile pod pojęciem statystyka rozumiemy architekta inferencji, czyli projektanta całego procesu wyciągania wniosków, a nie jedynie rzemieślnika od powtarzalnych rachunków. Ta druga funkcja, polegająca na manualnym mieleniu cyfr, była historycznie skazana na automatyzację od momentu, gdy w głowie pierwszego wizjonera zaświtała myśl o maszynie liczącej. Prawdziwa praca w tej dziedzinie nie polega dziś na samotnym generowaniu kolumn wyników, lecz na ustanawianiu warunków, pod którymi liczby w ogóle nabierają znaczenia. Projekt Zhana nie tylko nie spycha człowieka na margines, ale wręcz czyni go centralnym punktem systemu, bo to on decyduje o estymandzie, czyli wielkości, którą chcemy oszacować. To człowiek wybiera klasy metod, zakres symulacji oraz moment, w którym model staje się godny zaufania.

To nie jest rozpaczliwa defensywa starego porządku przed naporem algorytmów, lecz trzeźwe przypomnienie, gdzie właściwie przebiega nasza suwerenność metodologiczna. Współczesna

analityka cierpi na zdradliwą iluzję, wedle której każda poprawa skuteczności predykcyjnej jest tożsama ze wzrostem wartości naukowej samej procedury. To niebezpieczne uproszczenie, zwłaszcza w statystyce flirtującej z głębokim uczeniem, gdzie łatwo pomylić moc operacyjną modelu z jego prawomocnością inferencyjną. Ta pierwsza to czysta sprawność inżynierska: zdolność do wyłapywania subtelnych wzorców, poprawa błędu średniokwadratowego czy po prostu brutalne przyspieszenie obliczeń na klastrach. Prawomocność inferencyjna, czyli gwarancja, że wynik pozostaje metodologicznie obroniony i zachowuje kontrolę nad błędami statystycznymi, to jednak zupełnie inna kategoria. Statystyka nie utraciła rozumu. Utraciła niewinność. Bez tej świadomości łatwo wpuścić do laboratorium epistemologicznego oszusta w garniturze z TensorFlow.

Projekt Zhana zasługuje na obronę właśnie dlatego, że odmawia zatrzymania się na poziomie technicznej sprawności i uparcie dopytuje o fundamenty legitymizacji. Nie interesuje go triumf modelu jako efektownej sztuki inżynierskiej, lecz model jako legalny mieszkaniec republiki wnioskowania, podlegający jej surowym prawom i rygorom. To rozróżnienie staje się dziś bezcenne, ponieważ znaczna część dyskursu technologicznego zachowuje się tak, jakby każda poprawa trafności była już samoistnym zwycięstwem nauki. Niestety, w badaniach klinicznych czy w zaawansowanej inferencji, trafność pozbawiona ścisłej kontroli staje się raczej nowoczesną formą epistemicznej nieodpowiedzialności. Spór o relację między mocą testu a kontrolą błędu pierwszego rodzaju nie jest nudnym detalem dla teoretyków, lecz samym rdzeniem architektury poznania. To nie jest księgowość. To jest ekonomia polityczna wnioskowania.

Moc jest oczywiście uwodzicielska, bo obiecuje skuteczniejsze odrzucanie fałszywych hipotez i błyskawiczne wykrywanie efektów terapeutycznych przy optymalnym wykorzystaniu dostępnych danych. Jednak błąd pierwszego rodzaju, czyli ryzyko błędnego uznania nieistniejącego efektu za prawdziwy, stanowi naszą granicę cywilizacyjną, której przekroczyć nie wolno. Jeśli ten wskaźnik zaczyna puchnąć ponad nominalny poziom, procedura przestaje być odważna, a staje się po prostu epistemicznie rozrzutna. Rozdaje ona naukowe uznanie zbyt lekką ręką, co w badaniach nad lekami przekłada się na konkretne skutki medyczne i regulacyjne. Właśnie dlatego porównania z algorytmami takimi jak XGBoost czy Random Forest są w pracy Zhana tak kluczowe dla zrozumienia problemu. Czasem Random Forest przychodzi z lasem drzew tak okazałym, że ochroniarz słusznie pyta o zgodę budowlaną, zanim wpuści go do świata twardej nauki.

Zhan nie uprawia bałwochwalstwa wobec sieci neuronowych, lecz wskazuje, że w konkretnych zadaniach potrafią one lepiej połączyć wydajność z kontrolą nominalnego poziomu testu. Ta różnica nie ma charakteru kosmetycznego; to fundamentalny podział na narzędzia jedynie produktywne oraz te, które są ustrojowo bezpieczne dla nauki. Doskonale widać to w koncepcji pre-specyfikacji ilościowej, czyli praktyce polegającej na wcześniejszym wytrenowaniu modelu i

zamrożeniu jego parametrów, co pozwala na deterministyczne działanie na danych rzeczywistych. W dawnym porządku wystarczało zaufanie do badacza, który deklarował plan analizy, ale w świecie potężnych symulacji i metod bootstrapowych to już nie wystarcza. Ilościowa pre-specyfikacja przenosi cnotę z poziomu obyczaju na poziom architektury, chroniąc integralność badania przed pokusą manipulacji wynikami.

Rygor zamiast magii: Jak uwiarygodnić AI w nauce

Propozycja Zhana nie eliminuje wszystkich zagrożeń, ale znacząco zawęża obszar, w którym można po fakcie dopasowywać obliczenia do pożądanego rezultatu. Nie jest to jedynie praktyczna sztuczka obliczeniowa, lecz próba ustrojowego unowocześnienia rygoru badawczego, swoista konstytucjonalizacja procedury. Polega ona na budowaniu architektury badawczej, która narzuca dyscyplinę i ogranicza swobodę improwizacji na danych, zastępując naiwne zaufanie do cnoty badacza systemowymi zabezpieczeniami rzetelności. Współczesne dokumenty regulacyjne poruszają się dokładnie w tym kierunku, szukając twardych reguł tam, gdzie dotąd panowała analityczna dowolność. To nie jest księgowość; to jest ekonomia polityczna wnioskowania, w której każda decyzja obliczeniowa ma swoją cenę i swoje metodologiczne konsekwencje.

FDA w obowiązującym wytycznych dotyczących adaptacyjnych badań klinicznych wyraźnie podkreśla znaczenie planowania, prowadzenia, analizy i raportowania takich projektów w sposób, który zabezpiecza wiarygodność i interpretowalność wyników. Z kolei projekt ICH E20, udostępniony do konsultacji w 2025 roku, akcentuje potencjał adaptacyjnych projektów dla przyspieszenia rozwoju leków i bardziej efektywnej alokacji zasobów. Robi to jednak bez obniżania naukowych i regulacyjnych standardów, co w świecie goniącym za szybkim sukcesem brzmi niemal jak manifest konserwatywny. Oznacza to, że najwyższa instancja współczesnej praktyki regulacyjnej nie odrzuca innowacji obliczeniowej, lecz domaga się, aby była ona wpisana w dyscyplinę interpretacyjną i proceduralną. Nauka, podobnie jak konstytucja, bywa najbezpieczniejsza wtedy, gdy cnotę zastępuje instytucją.

Dokładnie to robi Zhan, broniąc nie samych sieci neuronowych, lecz ich użycia w ramach systemu, który zachowuje nienaruszalne własności graniczne. Warto wyostrzyć tę kategorię, czyli statystyczne gwarancje zachowywane nawet przy niedoskonałościach modelu lub utracie precyzji aproksymacji, bo to pojęcie kluczowe dla bezpieczeństwa wnioskowania. Jeśli model lub procedura zaczynają zawodzić, badanie musi mimo wszystko zachować pewne parametry, bez których przestaje być wiarygodnym instrumentem poznania. Kontrola błędu pierwszego rodzaju jest jedną z nich, niewielkie obciążenie estymatora drugą, a spójność zachowania przy zmianie scenariusza

trzecią. W dobrze zaprojektowanym układzie algorytmicznym można przegrać trochę lokalnej elegancji, byle nie przegrać konstytucji całości.

To brzmi niemal banalnie, lecz właśnie na tym polega różnica między inżynierią dla benchmarku a inżynierią dla nauki, gdzie ta pierwsza lubi błysk, a druga potrzebuje bezpieczników. Z tego punktu widzenia estymator zespołowy ma znaczenie niemal symboliczne, choć nie jest najbardziej romantycznym z możliwych rozwiązań. Nie obiecuje on całkowitej autonomii modelu ani nie sprzedaje marzenia, że sieć sama wydestyluje prawdę z surowych danych niczym matematyczny alchemik. Jego siła polega na czymś bardziej dojrzałym: łączy znane estymatory bazowe z DNN do uczenia wag, które poprawiają wynik końcowy. To rozwiązanie zachowuje więź z klasyczną inferencją i właśnie dlatego jest metodologicznie bardziej stabilne niż radykalne technologiczne zerwania.

W świecie technologicznym, który kocha narcystyczne architektury i radykalne manifesty, brzmi to może mało efektownie, ale nauka nie jest konkursem piękności dla algorytmów. Czasem najcenniejsza bywa metoda, która wie, skąd pochodzi, i nie wstydzi się własnych przodków, nawet jeśli noszą oni niemodne, statystyczne garnitury. Na tym tle bardzo ciekawe staje się zagadnienie interpretowalności z perspektywy biznesowej, gdzie odbiorca dojrzały nie potrzebuje naiwnych opowieści o intuicyjnych wykresach. Sektor farmaceutyczny czy biotechnologiczny chce wiedzieć, czy procedura da się obronić przed audytem i czy wyniki można zakomunikować w sposób odpowiedzialny. Interpretowalność w tym środowisku nie oznacza pełnej psychologicznej przejrzystości modelu, lecz zdolność do wbudowania go w porządek odpowiedzialności.

Właśnie dlatego rozdział Zhana o interpretowalnych modelach regresji, opartych na komponentach bazowych wynikających z wiedzy dziedzinowej, zasługuje na szczególną uwagę. Oferuje on model interpretowalności operacyjnej, czyli podejście skupione na praktycznej użyteczności i możliwości weryfikacji, zamiast na sentymentalnym dążeniu do prostoty. Taki model jest znacznie bliższy rzeczywistym potrzebom nauki i biznesu niż szkolne przeciwstawienie: proste versus czarne pudełko. Zhan pokazuje, że integralność badania, czyli sposób zaprojektowania procedur tak, aby wynik nie był podatny na manipulację po fakcie, jest ważniejsza niż estetyka kodu. Statystyka nie utraciła rozumu, utraciła niewinność, ale dzięki takim propozycjom zyskuje nową, cyfrową podmiotowość.

Infrastruktura jako konstytucja nowej statystyki

Warto przywołać perspektywę zarysowaną w tekstach o rewolucji równoległości, która uświadamia nam, że współczesna infrastruktura obliczeniowa nie jest jedynie technicznym zapleczem, lecz

ukrytą konstytucją nowoczesnej wiedzy. Infrastruktura obliczeniowa, czyli zasoby sprzętowe i moc procesorów graficznych traktowane jako fundament współczesnej metodologii, przesuwa punkt ciężkości z samego modelu na ustrój obliczeń i geopolitykę dostępu do narzędzi. Gdy jednostki GPU i przetwarzanie równoległe stają się warunkiem brzegowym wykonalności potężnych symulacji, metodologia nie może już udawać, że sprzęt pozostaje obojętny wobec epistemologii. To nie jest tylko kwestia szybszych procesorów, lecz fundamentalna zmiana w sposobie, w jaki konstruujemy dowody naukowe. Homer spotyka administratora klastra w serwerowni, gdzie krzem dyktuje nowe warunki teorii poznania.

Podobną rewolucję obserwujemy w wymiarze czasu, gdzie procedury wspomagane przez głębokie sieci neuronowe skracają czasochłonne zadania z wielu dni do zaledwie kilku minut. Taka zmiana nie służy wyłącznie wygodzie badacza, lecz radykalnie przekształca rozkład realnych możliwości poznawczych, pozwalając na znacznie częstsze walidacje i szersze badanie scenariuszy. Możemy dzięki temu szybciej reagować w badaniach adaptacyjnych i precyzyjniej dokumentować procesy decyzyjne, co wcześniej było blokowane przez barierę wydajności. Infrastruktura przyspieszająca procesy myślowe staje się zatem integralnym elementem nowego ładu racjonalności, a nie luksusowym dodatkiem do tradycyjnego warsztatu. Statystyka nie utraciła rozumu. Utraciła niewinność.

Współczesna recepcja nauki musi zatem operować na trzech poziomach jednocześnie, widząc model, metodologię oraz infrastrukturę jako nierozerwalną całość. Kto skupia wzrok wyłącznie na samym modelu, ten widzi jedynie maskę, za którą kryją się znacznie głębsze mechanizmy kształtujące wynik końcowy. Jeszcze wyraźniej problem ten ujawnia się w kontekście luk w danych, gdzie statystyka wspierana przez sztuczną inteligencję ryzykuje stanie się maszyną do powielania dawnych asymetrii. Jeśli nie zrozumiemy, że mechanizm generowania danych jest odbiciem społecznej historii pomiaru, algorytm zadziała jedynie jako potężny wzmacniacz strukturalnego błędu. Dane niereprezentatywne, obciążone niejawnie uprzywilejowanym punktem odniesienia, nie stają się neutralne tylko dlatego, że przepuścimy je przez wyszukany proces obliczeniowy.

Właśnie dlatego safeguarding, czyli systemowe zabezpieczanie rzetelności procesu, nie może ograniczać się do technicznej ochrony przed wartościami spoza dopuszczalnego zakresu geometrycznego. Musi on objąć również sferę społeczną, kliniczną i populacyjną, uwzględniając rzeczywiste profile ryzyka, które często nie mieszczą się w domyślnym modelu świata wpisanym w zbiory treningowe. Jeżeli realni pacjenci i ich zróżnicowane potrzeby zostaną pominięci w fazie projektowania, cała elegancja obliczeniowa pozostanie poznawczo podejrzana i etycznie ułomna. To uzupełnienie nie osłabia tez źródłowych, lecz wręcz je wzmacnia, pokazując, że ochrona integralności jest kategorią znacznie szerszą niż czysta technika.

Potrzebujemy dziś odwagi intelektualnej, by głośno powiedzieć, że statystyka przyszłości nie będzie ani wyłącznie klasyczna, ani czysto algorytmiczna. Podejście klasyczne, ignorujące złożoność współczesnych problemów, szybko zamieni się w muzeum elegancji, piękne, lecz całkowicie nieprzydatne w starciu z chaosem danych. Z kolei podejście czysto algorytmiczne, wierzące, że sama pojemność modelu rozstrzyga o jakości wiedzy, stanie się jedynie fabryką szybkich i kosztownych złudzeń. Przyszłość należy do systemów hybrydowych, w których głębokie uczenie pracuje jako komponent wyspecjalizowany, lecz zawsze podporządkowany rygorowi matematycznemu. W takich układach kluczowe są własności graniczne, czyli statystyczne gwarancje zachowywane nawet przy niedoskonałościach modelu lub utracie precyzji aproksymacji.

W tym nowym porządku kluczową rolę odgrywa pre-specyfikacja ilościowa, czyli praktyka polegająca na przeniesieniu kluczowych elementów zmienności z fazy po zebraniu danych do fazy przed badaniem poprzez sztywne reguły. Pozwala to ograniczyć uznaniowość analityka i chroni proces przed manipulacjami, które mogłyby dopasować wynik do oczekiwań politycznych czy biznesowych. Integralność badania, rozumiana jako sposób zaprojektowania procedur analitycznych tak, aby wynik nie był podatny na manipulację po fakcie, staje się gwarantem obiektywizmu. To nie jest kompromis wynikający z braku zasobów, lecz forma wyższej świadomości, która broni rozumu instytucjonalnego przed pokusami improwizacji. Pojawia się tu sens profetyczny całego przedsięwzięcia, który można sformułować bez zbędnej egzaltacji i patosu.

W najbliższych latach nie zwyciężą ci, którzy dysponują najgłośniejszymi modelami, lecz ci, którzy potrafią zamienić algorytm w procedurę godną zaufania. Sukces odniosą badacze budujący systemy zdolne do audytu i odporne na manipulację, a nie ci, którzy mylą automatyzację z suwerennością poznawczą. Nie wygrają również ci, którzy ignorują politykę danych i reprezentacji, gdyż ich eleganckie estymatory rozbiją się o rzeczywistość społeczną równie boleśnie jak o błędy parametryczne. Nowa statystyka nie będzie miała jednej, uniwersalnej twarzy, lecz będzie posiadała własną, surową konstytucję. Konstytucjonalizacja procedury, czyli budowanie architektury badawczej narzucającej dyscyplinę i ograniczającej swobodę improwizacji na danych, zastępuje zaufanie do cnoty badacza systemowymi zabezpieczeniami.

W tym tkwi najgłębszy paradoks współczesności, gdzie głębokie uczenie, oskarżane o zaciemnianie nauki, może stać się narzędziem jej większej uczciwości. Dzieje się tak nie dlatego, że sztuczna inteligencja jest magicznie mądrzejsza od człowieka, lecz dlatego, że pozwala nam ograniczyć własne słabości. Architekt inferencji, czyli nowoczesna rola statystyka jako projektanta problemu i strażnika nienegocjowalnych zasad wnioskowania, definiuje na nowo unikalną wartość ludzkiej inteligencji. To nie jest opowieść o detronizacji badacza, lecz o jego ostatecznym dojrzeniu do roli projektanta bezpiecznych systemów. Badacz, który rozumie wagę integralności, nie pragnie

nieskończonej swobody, lecz procedury, która powstrzyma go przed dopasowywaniem świata do wyniku.

Statystyka jako instytucja: Nowy rygor w dobie algorytmów

Statystyka, wbrew powszechnym wyobrażeniom, nie jest jedynie chłodnym zapisem faktów, lecz przede wszystkim surowym narzędziem dyscypliny intelektualnej. Jeżeli testowanie hipotez stanowi publiczną, nieco wyreżyserowaną twarz tej dziedziny, to estymacja punktowa pozostaje jej prawdziwym, nie zawsze fotogenicznym obliczem. To właśnie w tym obszarze rozstrzyga się, czy badacz pojmuje fundamentalną różnicę między spektaklem uzyskanego wyniku a odpowiedzialnością za jego wewnętrzną strukturę. W świecie adaptacyjnych badań klinicznych problem ten staje się wyjątkowo palący, gdyż dane przestają być spokojnym osadem z jednolitego mechanizmu losowania. Stają się one produktem dynamicznego procesu, w którym reguły przydziału i rozmiar próby ewoluują zgodnie z wcześniej przyjętym protokołem. W takim świecie naiwny estymator przestaje być niewinny, stając się po prostu przejawem metodologicznej lekkomyślności.

Propozycja Tianyu Zhana, aby zaprząc głębokie sieci neuronowe (DNN) do wspomagania tego procesu, nie jest jedynie technologiczną ciekawostką dla miłośników nowinek. To raczej przemyślana odpowiedź na wyzwania, których klasyczne rozwiązania analityczne nie potrafią udźwignąć w eleganckiej, zamkniętej formie matematycznej. Przewaga myślenia hybrydowego nad ślepym algorytmizmem staje się tutaj niemal fizycznie odczuwalna, oferując rygor tam, gdzie dotąd panowała niepewność. Estymator zespołowy wspierany przez DNN nie kusi obietnicą całkowitego zerwania z uświęconą tradycją statystyczną, co czyni go konstrukcją niezwykle dojrzałą. Jego atrakcyjność płynie z faktu, że zachowuje on stały kontakt z bazowymi estymatorami o doskonale znanych, bezpiecznych własnościach teoretycznych. Jednocześnie potrafi on wykorzystać sieć neuronową do precyzyjnego wyuczenia wag, które optymalizują kompromis między obciążeniem a błędem średniokwadratowym.

W nauce bowiem prawdziwe bezpieczeństwo nie wynika z tego, co wydaje się najbardziej imponujące w pojedynczym przebiegu eksperymentu. Prawdziwy rygor objawia się w tym, co pozostaje przewidywalne przy zmianie scenariusza i wyjściu poza komfortowy zakres danych. W tym specyficznym ujęciu estymator zespołowy zachowuje się jak rozsądny konstytucjonalista, podczas gdy estymator bezpośredni przypomina utalentowanego, lecz pozbawionego hamulców polityka. Pierwszy z nich bywa może mniej błyskotliwy w doraźnym działaniu, lecz znacznie częściej

okazuje się godny zaufania w dłuższej perspektywie. Ta przewaga nie ma wyłącznie charakteru matematycznego, gdyż posiada ona również bardzo konkretny, wręcz brutalny wymiar ekonomiczny. To nie jest zwykła księgowość, to jest ekonomia polityczna wnioskowania, gdzie każda cyfra niesie ze sobą ciężar realnych decyzji.

W skomplikowanych projektach rozwoju nowych leków błąd estymacji nigdy nie jest tylko abstrakcyjną skazą na tablicy w gabinecie profesora. Może on oznaczać fatalną decyzję inwestycyjną, błędną ocenę relacji korzyści do ryzyka lub tragiczne w skutkach opóźnienie w dostarczeniu ratującej życie terapii. Ekonomista doskonale rozumie, że wartość informacji nie zależy wyłącznie od jej formalnej poprawności, lecz od jej odporności na wstrząsy. Jeśli model wspierany przez DNN poprawia jakość estymacji przy zachowaniu właściwości granicznych, czyli gwarancji matematycznych zachowywanych nawet przy niedoskonałościach modelu, daje nam coś więcej niż liczbę. Dostarcza on solidniejszej podstawy do alokacji zasobów, co w sektorze klinicznym zawsze oznacza alokację szans i czasu życia realnych ludzi. Dlatego spór o architekturę estymatora nie jest nigdy wyłącznie jałową debatą akademicką, lecz walką o gospodarkę decyzji w warunkach niepewności.

Należy też otwarcie przyznać, że czas obliczeniowy przestał być jedynie technicznym zapleczem ukrytym za eleganckim salonem teorii naukowej. Stał się on jedną z kluczowych walut wiarygodności, organizującą hierarchie możliwości poznawczych we współczesnej nauce. Jeżeli procedura analityczna jest tak powolna, że uniemożliwia szerokie badanie scenariuszy lub szybkie symulacje, jej formalne piękno staje się bezużyteczne. Zhan trafnie przesuwa ciężar pracy na etap wcześniejszy, tworząc najpierw ogromny, sztucznie wygenerowany świat uczący dla modelu. Dzięki temu, gdy napływają realne dane, analiza przestaje być kosztowną improwizacją i staje się błyskawicznym wykonaniem precyzyjnie przygotowanej procedury. Ta logika jest jednocześnie obliczeniowa, regulacyjna i głęboko filozoficzna, ponieważ zastępuje doraźną intuicję trwałą architekturą systemu.

Tutaj właśnie rezonuje perspektywa obecna w tekstach o rewolucji równoległości, która pokazuje, że sztuczna inteligencja to nie tylko abstrakcyjne równania. To przede wszystkim infrastruktura mocy obliczeniowej, czyli zasoby sprzętowe traktowane jako fundament współczesnej metodologii, które determinują, kto może walidować swoje tezy. Kto dysponuje odpowiednią równoległością, ten może szybciej trenować modele i skuteczniej aktualizować procedury badawcze przed podjęciem kluczowych decyzji. Kto tej mocy nie posiada, bywa skazany na intelektualny outsourcing i niebezpieczne uproszczenia, które osłabiają rygor naukowy. Metoda statystyczna wspierana przez DNN jest więc nierozerwalnie zanurzona w polityce infrastruktury, a nie jest jedynie romantyczną

opowieścią o czystym rozumie. To historia o serwerowniach i akceleratorach, gdzie matematyczny arystokrata musi czasem ubrudzić ręce przy konfiguracji klastra obliczeniowego.

W tym punkcie statystyk niespodziewanie spotyka geopolitykę, uświadamiając sobie, że przezroczystość komputera jest równie iluzoryczna jak bezstronność nowoczesnego państwa. Infrastruktura obliczeniowa realnie decyduje o tym, kto może pozwolić sobie na rygorystyczną walidację, a kto musi iść na zgniłe kompromisy. Statystyka wspierana przez deep learning staje się zatem pytaniem o technologiczną suwerenność i ustrój techniki, w którym model nie jest narzędziem cudzej hegemonii. Nie wystarczy mieć genialny model, trzeba jeszcze posiadać warunki, w których nie stanie się on jedynie trybem w obcej maszynerii. W praktyce to właśnie dostęp do GPU rozstrzyga, czy dana wspólnota naukowa występuje w roli podmiotu, czy jedynie użytkownika cudzych narzędzi. To przesunięcie środka ciężkości ku kwestii ustroju techniki pozwala nam dostrzec głębszy sens współczesnych sporów metodologicznych.

Warto w tym miejscu powrócić do samej idei pre-specyfikacji, gdyż to ona odróżnia rzetelną propozycję od efektownego, lecz pustego triku. W klasycznej kulturze badań klinicznych pre-specyfikacja jakościowa polegała na opisaniu z góry planowanych działań, co było krokiem ważnym, lecz kruchym. Wiele zaawansowanych procedur pozostawiało analitykowi zbyt szerokie pole do interpretacji już po zobaczeniu realnych wyników, co otwierało drzwi dla metodologicznej arbitralności. Zhan proponuje jednak ruch znacznie ambitniejszy, wprowadzając pre-specyfikację ilościową, czyli przeniesienie kluczowych elementów zmienności do fazy przed badaniem poprzez sztywne reguły. Zamrożony model staje się deterministycznym mechanizmem wykonania, który strukturalnie ogranicza późniejszą uznaniowość badacza i chroni proces przed manipulacją typu p-hacking.

Taka konstrukcja idealnie współgra z duchem nowoczesnych wymogów regulacyjnych, gdzie nacisk kładzie się na interpretowalność i ochronę integralności badania. Integralność badania, czyli sposób zaprojektowania procedur tak, aby wynik nie był podatny na manipulację po fakcie, staje się tu fundamentem. Pre-specyfikacja ilościowa jest w istocie próbą przekształcenia statystyki z ulotnego etosu w trwałą i przewidywalną instytucję kontroli. To wielki krok cywilizacyjny, przypominający moment, w którym w sferze publicznej przestaliśmy liczyć wyłącznie na cnotę urzędnika. Zrozumieliśmy wówczas, że potrzebujemy systemów ograniczeń, rozdziału kompetencji i pełnej dokumentowalności podejmowanych działań. Statystyka nie utraciła rozumu, ale utraciła niewinność, zastępując naiwne zaufanie do badacza solidną architekturą nienegocjowalnych zasad wnioskowania.

Można wręcz zaryzykować stwierdzenie, że nowoczesny statystyk staje się architektem inferencji, czyli projektantem problemu i strażnikiem nienegocjowalnych zasad wnioskowania w świecie

automatów. Jego rola nie polega już na ręcznym liczeniu średnich, lecz na konstruowaniu systemów, które są odporne na ludzkie słabości i pokusy. W dobie algorytmów rygor nie płynie z samej matematyki, lecz z instytucjonalnej oprawy, w której ta matematyka jest osadzona. Każda linia kodu w zamrożonym modelu DNN staje się zapisem w cyfrowej konstytucji badania, której nie można zmienić w trakcie trwania procesu. A nauka, podobnie jak konstytucja, bywa najbezpieczniejsza wtedy, gdy cnotę zastępuje instytucją, budując architekturę nienegocjowalnych zasad wnioskowania. Dopiero wtedy statystyka kliniczna staje się prawdziwym systemem kontroli, a nie jedynie etyczną deklaracją bez pokrycia w infrastrukturze.

Architektura odpowiedzialności: Nowy ustrój obliczeniowej roztropności

Stara maksyma mówi, że badacz powinien być człowiekiem uczciwym, lecz współczesna nauka uczy nas, że nie powinien być on zmuszany do nieustannego dowodzenia tej cnoty. Statystyka nie utraciła rozumu, ale z pewnością utraciła niewinność, ponieważ swoboda improwizacji na danych, które już raz ujrzały światło dzienne, jest pokusą niemal niemożliwą do odparcia. Zamiast liczyć na hart ducha analityka, lepiej zbudować system, który tę improwizację po prostu technicznie uniemożliwi, stawiając rygor ponad intuicję. W tym ujęciu dobrze wytrenowana i zamrożona sieć neuronowa przestaje być czarną skrzynką, a staje się rodzajem cyfrowego kagańca nałożonego na badawczą dowolność. To rozwiązanie brzmi przewrotnie, gdyż kultura popularna przyzwyczaiła nas do widzenia w sztucznej inteligencji raczej żywiołu chaosu niż surowego narzędzia konstytucyjnej dyscypliny.

W odpowiednio zaprojektowanej metodzie głębokie sieci neuronowe mogą stać się fundamentem ograniczania interpretacyjnej anarchii, która tak często trawi współczesne nauki empiryczne. Bezpieczeństwo metody, rozumiane jako odporność na kaprysy świata rzeczywistego, staje się tutaj wartością nadrzędną i niemal ustrojową. Wartość projektu Tianyu Zhana ujawnia się najpełniej nie w momentach, gdy procesy przebiegają gładko, lecz wtedy, gdy rzeczywistość przestaje być uprzejma dla badacza. Problem out of range, czyli sytuacja, w której parametry procesu wychodzą poza wcześniej przygotowaną przestrzeń, jest pod tym względem jednym z najbardziej pouczających motywów całej pracy. Pokazuje on dobitnie, że model może błyszczeć na danych treningowych, by chwilę później gwałtownie stracić wiarygodność, gdy tylko realny świat postanowi nie składać przysięgi na nasze zakresy symulacyjne.

To nie jest marginalny błąd techniczny, lecz codzienna rzeczywistość nauki, która musi mierzyć się z nieprzewidywalnością materii. Dlatego safeguarding, czyli konstytucjonalizacja procedury

polegająca na budowaniu architektury badawczej narzucającej dyscyplinę, musi oznaczać coś znacznie więcej niż tylko dobre wyniki w warunkach laboratoryjnego komfortu. Prawdziwa odporność systemu objawia się w zdolności do przetrwania momentu, w którym świat okazuje się bardziej asymetryczny lub po prostu bardziej nieczysty, niż zakładał to pierwotny model uczący. W tym właśnie miejscu konstrukcje zespołowe, oparte na spójnych estymatorach bazowych i ograniczeniach strukturalnych, ujawniają swoją wyższość nad agresywnymi rozwiązaniami bezpośrednimi. Można to ująć jeszcze szerzej: błąd out of range nie dotyczy wyłącznie suchych parametrów liczbowych, lecz uderza w same fundamenty populacji i kontekstów społecznych.

Jeśli system analityczny został wzniesiony na fundamencie rzeczywistości zredukowanej, zdominowanej przez jedną grupę lub jeden standardowy wzorzec ciała, model staje się poznawczo kruchy. Luka w danych, o której tak trafnie pisze autor, wskazuje, że rzekoma neutralność statystyki bywa często jedynie przebranym uprzywilejowaniem męskiej normy. W medycynie takie przeoczenie nie jest jedynie błędem estetycznym w opisie, lecz realnym zagrożeniem klinicznym, które może kosztować ludzkie życie. W kontekście metod wspieranych przez sztuczną inteligencję wniosek płynący z tych rozważań jest brutalnie szczery i całkowicie bezlitosny. Nie da się uczciwie mówić o bezpieczeństwie procedur, jeśli chroni się jedynie matematyczny poziom istotności, ignorując jednocześnie reprezentacyjną uczciwość materiału wyjściowego.

W przeciwnym razie otrzymujemy jedynie bardzo elegancką, niezwykle kosztowną maszynę do reprodukowania historycznych asymetrii i uprzedzeń. To rozszerzenie definicji rygoru ma również głęboki sens antropologiczny, ponieważ każdy system ekspercki produkuje pewną wizję człowieka, nawet gdy milczy na ten temat. Badanie kliniczne z definicji ustala, czyje ciało jest wzorcem, jakie odchylenia są uchwytne, a jakie skutki uboczne uznaje się za typowe dla gatunku. Jeśli statystyka wspierana przez deep learning ma pretendować do roli nowoczesnej dyscypliny, nie może pozostać ślepa na społeczną historię własnych danych. Musi wbudować w swoje mechanizmy obronne czujność wobec ukrytych asymetrii populacyjnych, co nie jest wtrętem ideologicznym, lecz elementarną obroną jakości wiedzy.

Model uczony na zbyt wąsko rozumianym świecie nie staje się bardziej prawdziwy tylko dlatego, że dysponuje większą mocą obliczeniową. Staje się on jedynie bardziej sprawny w produkowaniu swoich własnych, cyfrowych iluzji, co czyni go epistemologicznym oszustem w garniturze z TensorFlow. Współczesna recepcja podobnych propozycji jest rozpięta między dwiema skrajnościami, które rzadko znajdują wspólny język. Z jednej strony mamy entuzjastów technologicznych, skłonnych widzieć w każdym sukcesie sieci neuronowych zapowiedź definitywnego końca dawnych sporów metodologicznych. Z drugiej strony okopał się obóz

konserwatywny, reagujący na modele głębokie tak, jakby każde ich użycie było automatycznym upadkiem interpretowalności i początkiem naukowej herezji.

Obie te postawy są intelektualnie wygodne, a przez to całkowicie niewystarczające do opisu złożoności problemu, przed którym stoimy. Projekt Zhana jest cenny właśnie dlatego, że nie daje taniej satysfakcji żadnej ze stron tego jałowego sporu. Nie ogłasza on końca statystyki klasycznej, ani nie kłania się modzie na nieograniczoną, niemal magiczną autonomię modelu. Przeciwnie, proponuje on układ republikański, w którym klasyczne zasady inferencji określają, co ma być chronione, a symulacja buduje świat treningowy. Sieć neuronowa aproksymuje to, czego teoria nie potrafi dostarczyć w praktycznie użytecznej formie, działając jako wyspecjalizowany organ wykonawczy.

Mechanizmy ograniczeń i walidacji pilnują, aby zysk obliczeniowy nie został kupiony kosztem utraty konstytucyjnych własności badania, takich jak obiektywizm czy powtarzalność. To nie jest kompromis połowiczny, lecz forma wyższej organizacji rozumu, w której technologia służy rygorowi, a nie go zastępuje. Z tej perspektywy najważniejsza staje się nie sama wydajność modelu mierzona w benchmarkach, lecz cała architektura odpowiedzialności. Musimy pytać, kto odpowiada za wybór wejść, kto za dobór zakresów symulacji i kto definiuje, które własności statystyczne są absolutnie nienegocjowalne. Te pytania są bardziej fundamentalne niż spór o to, czy dany model osiąga wynik lepszy o kilka punktów procentowych od swojego poprzednika.

Nauka dojrzała nie mierzy swojej jakości wyłącznie zwycięstwem w tabeli wyników, lecz zdolnością do utrzymania sensu w świecie regulacji i ryzyka. Właśnie tam, w obszarze statystyki klinicznej, błędy przestają być abstrakcją, a stają się konkretnym zagrożeniem dla ludzkiego zdrowia. Gdy zebrać wszystkie nici tego wywodu i związać je bez zbędnego sentymentalizmu, otrzymujemy obraz wyjątkowo wyraźny i spójny. Współczesne metody statystyczne wspierane przez deep learning nie są ani prostą kontynuacją dawnej inferencji, ani jej gwałtownym zaprzeczeniem. Są one raczej próbą zbudowania nowego ustroju obliczeniowej roztropności, w którym technologia zna swoje miejsce w hierarchii wartości.

W tym nowym ustroju sieć neuronowa nie pełni funkcji monarchy, nie jest źródłem normy ani suwerenem ostatecznej decyzji badawczej. Jest ona raczej wyspecjalizowanym urzędnikiem, któremu wolno bardzo wiele, pod warunkiem, że najważniejsze zasady zostały zdefiniowane gdzie indziej. Zostały one określone przez projekt badania, przez wymóg kontroli błędu oraz przez obowiązek zachowania właściwości brzegowych, czyli statystycznych gwarancji zachowywanych nawet przy niedoskonałościach modelu. Taki jest rzeczywisty sens propozycji Tianyu Zhana i dlatego ta wizja zasługuje na merytoryczną obronę przed obiema skrajnościami. Nie składa ona

statystyki w ofierze sztucznej inteligencji, lecz raczej każe tej ostatniej zdać egzamin z posłuszeństwa wobec twardych reguł nauki.

To właśnie odróżnia poważną metodologię od technologicznej dewocji, która pyta jedynie o to, czy model „działa" w danym momencie. Prawdziwa nauka pyta, w jakich warunkach model przestaje działać, co dokładnie optymalizuje i jakie ryzyko wprowadza do systemu wiedzy. Deep learning assisted statistics, jeśli rozumieć ją zgodnie z duchem tej książki, odpowiada właśnie na te najtrudniejsze i najbardziej niewygodne pytania. Przyznaje ona otwarcie, że modele głębokie mają przewagę tam, gdzie klasyczne rozwiązania analityczne są zbyt kosztowne lub zbyt kruche. Jednocześnie ostrzega, że bez nadzoru eksperckiego i bez matematycznych ograniczeń, modele te mogą stać się źródłem błędów niebezpiecznych właśnie dlatego, że są tak błyskotliwe.

To trzeźwość bardzo rzadka w epoce, która kocha rozwiązania totalne i proste odpowiedzi na złożone problemy współczesności. W gruncie rzeczy spór dotyczy dziś nie tylko techniki, lecz także całego cywilizacyjnego stylu używania rozumu w służbie prawdy. Jedna szkoła myślenia chciałaby zachować klasyczną statystykę jako niemal nietknięte sanktuarium elegancji, w którym każda komplikacja świata musi cierpliwie czekać na teorię. Druga szkoła woli pójść w przeciwną skrajność i uznać, że wystarczy odpowiednio szybki hardware, by prawda sama wyłoniła się z procesu optymalizacji. A nauka, podobnie jak konstytucja, bywa najbezpieczniejsza wtedy, gdy cnotę zastępuje instytucją, a entuzjazm – architekturą odpowiedzialności. To nie jest zwykła księgowość, to jest polityczna ekonomia wnioskowania, w której stawką jest integralność naszego poznania.

Statystyka republikańska: między mocą obliczeń a rygorem metod

Obie dominujące dziś postawy wobec danych grzeszą słabością, choć każda na swój specyficzny sposób. Pierwsza z nich, zakorzeniona w metodologicznym konserwatyzmie, zdaje się całkowicie ignorować skalę oraz oszałamiającą złożoność problemów, przed którymi staje współczesna biostatystyka. Druga natomiast, zachłyśnięta czystą mocą obliczeniową, zapomina, że optymalizacja sama w sobie nie posiada kompasu moralnego ani poznawczego. Nie zna ona pojęcia dobra naukowego, nie dba o integralność badania, czyli sposób zaprojektowania procedur tak, aby wynik nie był podatny na manipulację po fakcie, ani nie rozumie ciężaru regulacyjnej odpowiedzialności. Przyszłość, jak sądzę, należy do trzeciej drogi – do systemów hybrydowych, w których potężna moc obliczeniowa pracuje w karbach architektury metodologicznej, a nie zamiast niej. To nie jest jedynie techniczna korekta, lecz fundamentalna zmiana paradygmatu, w którym statystyka nie utraciła rozumu, ale z pewnością utraciła niewinność.

Właśnie dlatego pojęcia takie jak prespecyfikacja ilościowa, czyli praktyka polegająca na przeniesieniu kluczowych elementów zmienności do fazy przed badaniem poprzez sztywne reguły, stają się dziś ważniejsze niż sam zachwyt nad głębokimi sieciami neuronowymi. W tym nowym układzie sił safeguarding oraz projektowanie scenariuszy walidacyjnych przestają być nudnym obowiązkiem biurokratycznym, a stają się warunkiem możliwości rzetelnego poznania. Szczególne znaczenie ma tutaj adaptacyjne badanie kliniczne, ponieważ to właśnie ono najdobitniej ujawnia pełną stawkę tego metodologicznego zwrotu. Adaptacyjność jest naukowo niezwykle atrakcyjna, gdyż pozwala wcześniej rozpoznać sygnał skuteczności leku, lepiej alokować pacjentów do grup i racjonalniej gospodarować ograniczoną próbą badawczą. Jest ona zarazem niebezpieczna, bo każda niekontrolowana zmiana w trakcie trwania eksperymentu stwarza ryzyko deformacji inferencji, jeśli nie została należycie zabezpieczona przed błędem systematycznym.

Współczesne instytucje regulacyjne, wbrew pozorom, wcale nie odrzucają tej nowoczesnej logiki, lecz starają się ją ująć w ramy cywilizowanego porządku. Amerykańska FDA w swoim guidance z 2019 roku oraz wytyczne ICH E20 w wersji z czerwca 2025 roku akceptują wartość projektów adaptacyjnych, o ile ich planowanie pozostaje rygorystycznie określone. Wymagają one, aby analiza i interpretacja wyników były odporne na uznaniowość, a wnioski końcowe pozostawały w pełni wiarygodne dla oceny relacji korzyści do ryzyka. To jest właśnie ten klimat intelektualny, w którym propozycja Zhana okazuje się nie być technologiczną ekstrawagancją, lecz precyzyjnym narzędziem wpisującym się w paradygmat nauki regulacyjnej. Można odnieść wrażenie, że Arystoteles patrzy tu na logi błędów, próbując odnaleźć w nich ślady dawnej logiki i porządku.

Nie wolno nam również zapominać, że czas obliczeniowy jest w tej historii kategorią znacznie głębszą, niż chcieliby to przyznać akademiccy puryści. Kto lekceważy realny koszt obliczeń, ten często nie rozumie, że praktyczny zakres symulacji oraz szybkość walidacji są integralną częścią samej jakości procesu poznawczego. Możliwość przeprowadzania błyskawicznych analiz wrażliwości i terminowość podejmowanych decyzji to nie są kwestie techniczne, lecz elementy składowe rzetelności naukowej. Zhan trafnie przenosi ciężar najtrudniejszej pracy na etap wcześniejszy, wykorzystując sztucznie generowane dane do treningu modelu i zamrożenia jego parametrów przed właściwym testem. Dzięki temu późniejsza analiza danych rzeczywistych może stać się procesem szybkim, deterministycznym i wolnym od pokusy dopasowywania teorii do wyników.

To nie jest jedynie kwestia wygody czy oszczędności zasobów, lecz specyficzna forma racjonalizacji całego procesu badawczego. Takie podejście drastycznie zmniejsza pole uznaniowości badacza po tym, jak zobaczy on już zgromadzone dane, co jest kluczowe dla zachowania obiektywizmu. W realiach współczesnej medycyny oznacza to skrócenie drogi między zebraniem surowej informacji a

podjęciem odpowiedzialnej decyzji o dopuszczeniu terapii. A w tej dziedzinie skrócenie wspomnianej drogi bywa różnicą nie między elegancją a brzydotą dowodu, lecz między realną stratą a pożytkiem dla pacjenta. Tutaj niezwykle trafne okazują się intuicje obecne na portalu Dobre Państwo, które przypominają nam o materialnym podłożu naszych intelektualnych ambicji.

Tekst o rewolucji równoległości przypomina, że GPU, CUDA czy TensorRT nie są już zwykłymi narzędziami technicznymi w rękach inżynierów. Stanowią one infrastrukturę obliczeniową, czyli fundament współczesnej metodologii oparty na zasobach sprzętowych, który współokreśla warunki możliwości dzisiejszej nauki i gospodarki. To bardzo ważne uzupełnienie dla refleksji nad statystyką wspieraną przez deep learning, ponieważ zdejmuje z niej maskę czystej abstrakcji. Oznacza to bowiem, że pytanie o metodę nie może być już oddzielone od pytania o materialne i geopolityczne warunki jej wykonalności. To nie jest księgowość, to jest polityczna ekonomia wnioskowania, w której dostęp do krzemu decyduje o jakości stawianych hipotez.

Kto posiada dostęp do potężnej, równoległej mocy obliczeniowej, ten może szerzej symulować, szybciej trenować modele i sprawniej wdrażać zaawansowane procedury hybrydowe. Kto tego dostępu jest pozbawiony, temu pozostaje albo bolesne uproszczenie metod, albo upokarzająca zależność od cudzej, często nieprzejrzystej infrastruktury. Współczesna biostatystyka spotyka się więc nie tylko z matematyką i medycyną, lecz także z twardą ekonomią polityczną techniki. Zhan ujmuje to zjawisko jako przejście od pojedynczego komponentu do złożonej, państwowej niemal infrastruktury. Skoro zaś ta infrastruktura współdecyduje o możliwościach naszego poznania, to nie wolno już mówić o sztucznej inteligencji jak o zjawisku bezpaństwowym i bezkosztowym.

Równie istotne jest drugie ostrzeżenie płynące z analiz dotyczących luki w danych, która zniekształca nasz obraz rzeczywistości. Świat zbudowany na danych pochodzących głównie od mężczyzn, przedstawiający ich jako normę uniwersalną, nie jest światem neutralnym, lecz statystycznie skrzywionym. Ma to bezpośrednie i często dramatyczne konsekwencje dla metod wspieranych przez algorytmy sztucznej inteligencji, które chłoną te uprzedzenia. Jeżeli model uczy się na materiale, który od samego początku niósł w sobie asymetrię reprezentacji, to będzie tę asymetrię utrwalał z jeszcze większą, nieludzką sprawnością. W takim świecie statystyka musi stać się czujna, by nie stać się jedynie narzędziem reprodukcji dawnych, systemowych wykluczeń.

Dlatego właśnie safeguarding nie może ograniczać się wyłącznie do matematycznej ochrony poziomu istotności alfa czy redukcji błędu średniokwadratowego. Musi on obejmować także głęboką czujność wobec tego, czy mechanizm generowania danych nie powiela cichego wykluczenia konkretnych populacji lub profili ryzyka. To jest ten moment, w którym antropologia i biostatystyka przestają być odległymi krewnymi i stają się najbliższymi współpracownikami w walce o prawdę. Dobra procedura statystyczna nie tylko liczy poprawnie, ale przede wszystkim wie,

kogo i co naprawdę reprezentuje w swoich tabelach. Z tej perspektywy rola human intelligence, czyli ludzkiego nadzoru nad procesem, staje się jeszcze bardziej centralna i nienegocjowalna.

W tym nowym układzie ekspert nie jest już tylko asystentem modelu, lecz to model staje się asystentem eksperta. To człowiek występuje jako architekt inferencji, czyli nowoczesny projektant problemu i strażnik nienegocjowalnych zasad wnioskowania, który odróżnia istotne parametry od szumu. To on decyduje, które ograniczenia estymatora są konieczne, rozpoznaje ryzyka przy błędnym mechanizmie generowania danych i wie, kiedy należy powiedzieć „stop". Pewne własności badania muszą pozostać nienaruszalne nawet wtedy, gdy najbardziej obiecujące symulacje podpowiadają badaczowi bardziej brawurowe i ryzykowne rozwiązania. Jeżeli współczesna kultura tak chętnie opowiada o całkowitej automatyzacji, to właśnie dlatego trzeba z uporem przypominać o suwerenności ludzkiego rozumu.

W nauce wysokiej stawki prawdziwa władza nie leży po stronie algorytmu, lecz po stronie tego, kto potrafi ustanowić reguły jego użycia. Bez tego algorytm nie jest narzędziem postępu naukowego, lecz jedynie szybszą i bardziej wyrafinowaną formą dawnej, ludzkiej uznaniowości. Można więc sformułować tezę, że bezpieczeństwo systemów AI w statystyce nie wynika z mitycznej doskonałości samego kodu. Wynika ono z siły porządku metodologicznego, który ten kod ogranicza, ukierunkowuje i osadza w procedurze odpornej na kaprysy świata. Właśnie dlatego najcenniejszym wkładem propozycji Zhana nie jest obietnica szybkości, lecz możliwość budowy procedur bardziej zdyscyplinowanych i lepiej chroniących integralność nauki.

To przejście od czystej, brutalnej skuteczności do czegoś, co moglibyśmy nazwać skutecznością konstytucyjną, jest prawdziwą nowością w naszej dziedzinie. Statystyka przyszłości nie będzie polegała na tym, że model bezbłędnie odgadnie stan świata, bo to jest mrzonka godna epistemologicznego oszusta w garniturze z TensorFlow. Będzie ona polegała na tym, że świat zostanie poddany analizie przez system, który nawet w warunkach błędu nie zdradza swoich fundamentalnych zasad. A nauka, podobnie jak konstytucja, bywa najbezpieczniejsza wtedy, gdy cnotę badacza zastępuje solidną, sprawdzoną instytucją. W tym sensie statystyka staje się formą ustroju poznania, w którym własności graniczne gwarantują bezpieczeństwo nawet przy niedoskonałej aparaturze.

Współczesny badacz stoi dziś pomiędzy trzema pokusami, niczym bohater antycznej tragedii zagubiony w cyfrowym świecie. Pierwsza z nich szepcze mu do ucha, by ufał wyłącznie dawnym, sprawdzonym metodom, choć świat i dane dawno już je przerosły. Druga obiecuje mu, że jeśli tylko uruchomi dość dużą sieć neuronową i nakarmi ją oceanem danych, to wszystkie problemy rozwiążą się same. Trzecia pokusa, paradoksalnie najrozsądniejsza, mówi mu coś znacznie mniej

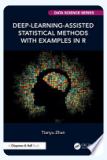
romantycznego i znacznie bardziej wymagającego. Każe mu ona budować procedury tak, jak buduje się mosty, a nie jak opowiada się piękne, lecz niemożliwe do zweryfikowania baśnie.

Ta trzecia szkoła nakazuje sprawdzać nośność, projektować marginesy bezpieczeństwa i zawsze zakładać możliwość wystąpienia awarii systemu. Uczy, by nie mylić estetycznej elegancji modelu z jego realną odpornością na błędy, ani szybkości obliczeń z ich prawną i naukową prawomocnością. Nigdy nie wolno ufać światu na tyle, by nie przygotować się na jego wrodzoną złośliwość i nieprzewidywalność. Właśnie tę szkołę reprezentuje najlepsza wersja statystyki wspieranej przez deep learning – szkoła bez kompleksów wobec nowej mocy, ale i bez nabożnego lęku. Rozumie ona, że myślenie statystyczne musi zawsze wyprzedzać moc obliczeniową, bo inaczej ta moc stanie się tylko głośniejszą formą błędu.

I to jest właściwa pointa naszych rozważań nad przyszłością tej dyscypliny. Nie potrzebujemy statystyki, która zamieni się w zakurzoną kaplicę starej, nieadekwatnej już elegancji, ani takiej, która stanie się kasynem szybkich aproksymacji. Potrzebujemy statystyki republikańskiej – instytucjonalnie mądrej, obliczeniowo śmiałej i konstytucyjnie ostrożnej w swoich sądach. Takiej, która przyjmuje deep learning do służby jako sprawnego pomocnika, ale nigdy nie oddaje mu suwerennego tronu decyzji. Takiej, która rozumie wartość infrastruktury, ale nie myli jej z ostateczną racją, i która walczy o moc testu, nie sprzedając za nią kontroli nad błędem. Prawdziwa konstytucjonalizacja procedury zaczyna się bowiem w rozumie, który potrafi wyznaczyć granice własnym narzędziom. Tianyu Zhan w swojej pracy o metodach statystycznych wspieranych przez deep learning daje nam do tego solidne podstawy.

> Statystyka staje się dziś nowym ustrojem poznania, w którym to nie algorytm jest suwerenem, lecz architektura odpowiedzialności, którą wokół niego wznosimy. Pytanie brzmi, czy potrafimy okiełznać moc obliczeniową tak, by zamiast szybkich złudzeń, dostarczała nam bezpiecznych gwarancji rzetelności. Czy staniemy się strażnikami własnych procedur, czy jedynie widzami w spektaklu, w którym maszyny piszą historię nauki za nas?

Inspiracje

[1] 

"Deep Learning assisted statistical methods with examples in R" Tianyu Zhan

2026 | Chapman and Hall/CRC | ISBN: 9781041158431

Tianyu Zhan jest dyrektorem w AbbVie Inc., gdzie koncentruje się na innowacyjnych projektach badań klinicznych i zaawansowanych metodach analizy statystycznej. W 2017 roku uzyskał tytuł doktora biostatystyki na Uniwersytecie Michigan w Ann Arbor. Jego praca zawodowa i zainteresowania badawcze koncentrują się na badaniach klinicznych w późnej fazie, wykorzystujących podejścia obliczeniowe do optymalizacji projektów i wyników badań. Zhan wniósł znaczący wkład w rozwój tej dziedziny, integrując techniki głębokiego uczenia z tradycyjnymi metodami statystycznymi, aby sprostać złożonym wyzwaniom w zakresie testowania hipotez, estymacji punktowej i optymalizacji. Jego praca ma na celu zwiększenie efektywności statystycznej i interpretowalności w zastosowaniach praktycznych, szczególnie w przemyśle biofarmaceutycznym. Jest doceniany za swoje wysiłki w promowaniu odpowiedzialnego i trafnego zastosowania metodologii opartych na sztucznej inteligencji w badaniach naukowych.

Tianyu Zhan *is a Director at AbbVie Inc., where he focuses on innovative clinical trial designs and advanced statistical analysis methods. He earned his Ph.D. in Biostatistics from the University of Michigan, Ann Arbor, in 2017. His professional work and research interests center on late-phase clinical trials, leveraging computational approaches to optimize trial design and outcomes. Zhan has contributed significantly to the field by integrating deep learning techniques with traditional statistical methods to address complex challenges in hypothesis testing, point estimation, and optimization. His work aims to enhance statistical efficiency and interpretability in practical applications, particularly within the biopharmaceutical industry. He is recognized for his efforts in promoting responsible and valid applications of AI-based methodologies in scientific research.*

AbbVie Inc.

Treść tworzy Fundacja Dobre Państwo.
Redaguje i publikuje APA ONE, autorski system redakcyjny Fundacji oparty na AI.
APA ONE Publishing System
Fundacja Dobre Państwo © 2026
Article ID: 5768f4d2-c606-471d-93e3-eed20a3ee436