# AI w obronności: Architektura selekcji i pułapki pewności


**AUTOR**

Fundacja Dobre Państwo



## STRESZCZENIE / ABSTRACT

Artykuł podejmuje krytyczną analizę roli sztucznej inteligencji w nowoczesnych systemach obronnych i sensorowych. Autor odrzuca uproszczone metafory AI jako autonomicznego podmiotu lub zwykłego narzędzia, definiując ją jako „architekturę selekcji" – reżim filtrujący rzeczywistość operacyjną. Kluczowym problemem jest zjawisko automation bias, czyli bezkrytyczne zaufanie operatorów do algorytmów, co prowadzi do pozornej kontroli nad systemem. Tekst argumentuje, że odpowiedzialność jest niezbywalną domeną człowieka, której nie można wyoutsourcować na maszynę. Wprowadzenie standardu Meaningful Human Control jest niezbędne, aby uniknąć redukcji człowieka do roli ceremonialnego świadka. Analiza obejmuje również techniczne aspekty, takie jak xAI, wywiad elektroniczny ELINT oraz algorytmy Q-RAM, osadzając je w kontekście etyki i ekonomii percepcji na polu walki.

---

This article critically examines the role of artificial intelligence in modern defense and sensor systems. The author rejects simplistic metaphors of AI as an autonomous entity or a mere tool, defining it as a "selection architecture"—a regime that filters operational reality. A key problem is the phenomenon of automation bias, or the uncritical trust of operators in algorithms, which leads to a false sense of control over the system. The article argues that responsibility is an inalienable human domain that cannot be outsourced to a machine. The introduction of the Meaningful Human Control standard is essential to avoid reducing humans to the role of a ceremonial witness. The analysis also covers technical aspects such as xAI, ELINT, and Q-RAM algorithms, placing them in the context of ethics and the economics of perception on the battlefield.


Współczesna sztuczna inteligencja w systemach sensorowych i obronnych przestała być jedynie narzędziem – stała się reżimem filtrowania rzeczywistości, który brutalnie narzuca ramy ludzkiemu rozumieniu pola walki. Artykuł dekonstruuje mit AI jako

autonomicznego bytu, wskazując, że w rzeczywistości jest ona architekturą selekcji, która redukuje złożoność świata kosztem selektywnego pomijania danych. Autor stawia tezę, że największym zagrożeniem nie jest błędne działanie maszyn, lecz automatyzacja podszywająca się pod kontrolę oraz cyniczne rozmywanie odpowiedzialności instytucjonalnej. W obliczu tzw. automation bias, człowiek w pętli decyzyjnej często staje się jedynie ceremonialnym świadkiem procesów, których już nie rozumie. Tekst dowodzi, że etyka w projektowaniu systemów wysokiego ryzyka nie jest estetycznym dodatkiem, lecz rygorystyczną teorią zarządzania niepewnością. Prawdziwa przewaga operacyjna nie wynika z gromadzenia danych, lecz z jakości modeli, które potrafią przyznać się do niewiedzy. Odpowiedzialność pozostaje domeną wyłącznie ludzką, a technologie takie jak xAI czy Meaningful Human Control powinny służyć nie jako alibi, lecz jako twarde narzędzia zachowania suwerenności decyzji w obliczu technologicznego determinizmu.

**SŁOWA KLUCZOWE**

architektura selekcji | automation bias | Meaningful Human Control | wyjaśnialna sztuczna inteligencja | xAI | systemy sensorowe | wywiad elektroniczny ELINT | klasyfikacja akustyczna | radary zwinne | open set recognition | algorytmy Q-RAM | Quality of Service QoS | mapy istotności | human in the loop | reżim filtrowania rzeczywistości

## AI jako reżim filtrowania rzeczywistości i pułapka odpowiedzialności

Refleksja nad obecnością sztucznej inteligencji w systemach sensorowych, obronnych i analitycznych cierpi dziś na osobliwą, niemal chroniczną chorobę języka. Z jednej strony serwuje się nam opowieści o autonomii, które brzmią jak zapowiedź metafizycznej emancypacji maszyny, z drugiej zaś próbuje się sprowadzić AI do roli narzędzia, jakbyśmy mieli do czynienia jedynie z nieco bardziej wyrafinowanym młotkiem. Obie te metafory są jednak poznawczo jałowe i niebezpiecznie nieadekwatne do rzeczywistości, z którą mierzą się operatorzy. W praktyce operacyjnej sztuczna inteligencja jest bowiem przede wszystkim architekturą selekcji, czyli systemem aktywnie filtrującym sygnały, hipotezy i priorytety, zanim w ogóle dotrą one do ludzkiej świadomości. To ona

decyduje, co w gąszczu danych zasługuje na uwagę, a co zostanie skazane na cyfrowy niebyt. W tym sensie AI nie jest ani osobą, ani zwykłym instrumentem, lecz reżimem filtrowania rzeczywistości, który brutalnie narzuca ramy naszemu rozumieniu pola walki.

To właśnie z tego powodu etyka sztucznej inteligencji nie może być traktowana jako estetyczny dodatek PR-owy do twardej inżynierii ani jako kojąca opowieść dla zarządów i ministerstw. Musi ona stać się rygorystyczną teorią odpowiedzialnego porządkowania mocy obliczeniowej, poznawczej i decyzyjnej w warunkach skrajnej niepewności. Kto nie potrafi dostrzec tej różnicy, ten myli głęboką transformację epistemiczną z prymitywnym marketingiem technologii, który karmi się obietnicami bez pokrycia. A marketing, jak wiadomo, potrafi sprzedać nawet mgłę jako infrastrukturę krytyczną, byle tylko nadać jej efektowny pulpit nawigacyjny i kilka obiecujących wykresów. Prawdziwe wyzwanie leży jednak gdzie indziej – w zrozumieniu, że każda redukcja złożoności, czyli proces upraszczania świata przez algorytm w celu umożliwienia szybkiej reakcji, jest w istocie aktem politycznym i moralnym.

W książce najcenniejsza wydaje się teza, że odpowiedzialność pozostaje domeną wyłącznie ludzką i nie podlega żadnej formie outsourcingu na rzecz maszyn. Nie wynika to bynajmniej z faktu, że człowiek jest istotą biologicznie sentymentalną czy kruchą, lecz z fundamentalnej natury samego pojęcia winy i kary. Odpowiedzialność nie jest bowiem własnością obliczeniową, którą można zoptymalizować, lecz statusem normatywnym, nierozerwalnie związanym z podmiotowością. Algorytm może z niezwykłą precyzją minimalizować funkcję kosztu, szacować prawdopodobieństwo wystąpienia zdarzenia czy wyłuskiwać subtelne wzorce z szumu informacyjnego. Może nawet produkować niezwykle przekonujące uzasadnienia dla własnych rekomendacji, imitując proces logicznego wnioskowania.

Nie może jednak w sensie ścisłym odpowiadać wobec kogokolwiek, ponieważ nie posiada sprawstwa moralnego ani zdolności do odczuwania ciężaru zobowiązania. Maszyna nie zna wstydu, nie podlega rygorom sumienia i nie posiada tego, co prawnicy nazywają dojrzałą zdolnością do ponoszenia konsekwencji własnych czynów. Właśnie dlatego antropomorfizacja AI jest nie tylko błędem intelektualnym, ale przede wszystkim cynicznym mechanizmem rozmywania odpowiedzialności instytucjonalnej. Książka wskazuje, że AI należy traktować wyłącznie jako produkt i narzędzie, jeśli chcemy, by odpowiedzialność twórców i operatorów pozostała egzekwowalna. Ta intuicja współbrzmi z regulacjami unijnymi oraz amerykańską debatą o bezpieczeństwie, gdzie akcentuje się minimalizowanie błędów zamiast przyznawania maszynom quasi-podmiotowości.

Stąd centralne pytanie etyki AI brzmi zupełnie inaczej, niż chcieliby tego futuryści karmiący nas tanim zachwytem nad cyfrową świadomością. Pytanie to jest znacznie ostrzejsze: czy wdrożenie

algorytmów realnie zwiększa ludzkie sprawstwo, czy też redukuje człowieka do roli ceremonialnego świadka procesów, których już nie ogarnia? W systemach bojowych i wywiadowczych prawdziwym przeciwnikiem nie jest sama automatyzacja, lecz automatyzacja podszywająca się pod kontrolę. Operator może bowiem pozostawać w pętli decyzyjnej tylko nominalnie, stając się jedynie biologiczną pieczęcią na wyrokach wydanych przez kod. Jeśli rekomendacja systemu jest zbyt szybka i zbyt złożona, by można ją było realnie zakwestionować w czasie rzeczywistym, zgoda człowieka staje się pustym rytuałem.

Tę pułapkę opisuje pojęcie automation bias, czyli tendencja operatorów do bezkrytycznego ufania wynikom generowanym przez algorytmy, nawet gdy są one ewidentnie błędne. W takim układzie człowiek przestaje kontrolować system, a system zaczyna kontrolować zakres tego, co człowiek może jeszcze sensownie zanegować. Aby temu zapobiec, konieczne jest wdrożenie standardu Meaningful Human Control, czyli nadzoru, w którym człowiek posiada realną, a nie tylko liturgiczną możliwość interwencji. Nie chodzi tu o fizyczną obecność przycisku „stop", lecz o realną zdolność rozumienia, przerwania i odrzucenia decyzji maszyny. W przeciwnym razie human in the loop staje się liturgią odpowiedzialności bez odpowiedzialności, gdzie winnego szuka się w kodzie, a nie w gabinetach.

Właśnie w tym kontekście pojawia się koncepcja wyjaśnialnej sztucznej inteligencji, znanej szerzej jako xAI, która nie jest luksusem dla akademików, lecz warunkiem operacyjnej trzeźwości. W zastosowaniach typu ATR, czyli systemach automatycznego rozpoznawania celów, wizualizacje takie jak mapy ciepła pełnią funkcję krytycznego punktu zaczepienia dla operatora. Ich zadaniem jest uchylenie wieka czarnej skrzynki, aby człowiek nie stał się ślepym notariuszem rekomendacji modelu, lecz świadomym uczestnikiem procesu. Programy badawcze, takie jak DARPA XAI, od lat podkreślają, że sukces uczenia maszynowego ogranicza nie tylko trafność, ale przede wszystkim niezdolność systemu do przedstawienia zrozumiałych powodów działania. Bez tej przejrzystości każda innowacja staje się jedynie nowym rodzajem poznawczego więzienia.

Równocześnie musimy zachować zdrowy sceptycyzm, gdyż xAI nie jest magicznym rozwiązaniem wszystkich problemów etycznych i technicznych. Komisja Europejska słusznie zauważa, że mapy istotności mogą wskazać, które dane miały znaczenie dla modelu, ale dają tylko ograniczony wgląd w to, jak zostały one faktycznie przetworzone. Innymi słowy, xAI nie jest oknem na duszę modelu, lecz raczej latarką skierowaną w stronę mechanizmu, który nadal pozostaje częściowo zacieniony. To już dużo, ale zbyt mało, by popaść w technologiczną dewocję i uznać problem za rozwiązany. Z punktu widzenia teorii organizacji xAI jest przede wszystkim instrumentem obniżania kosztów transakcyjnych odpowiedzialności, pozwalającym na jaśniejszy podział kompetencji między ludźmi a systemami.

Im mniej przejrzysty jest system, tym trudniej rozdzielić kontrolę i winę w sytuacjach kryzysowych, co w organizacjach wysokiego ryzyka generuje ogromne koszty audytu i sporów prawnych. Transparentność modelu działa więc jak mechanizm ograniczania hazardu moralnego między projektantem, operatorem a instytucją zamawiającą dany system. Fundacja Dobre Państwo, pisząc o architekturze innowacji, trafnie podkreślała rolę guardrails oraz potrzebę ochrony przed tak zwanym teatrem KPI, czyli pogonią za wskaźnikami, które nie oddają realnej jakości procesów. Ta intuicja ma bezpośrednie zastosowanie w systemach obronnych, gdzie aksjologia progu alarmowego, czyli wybór czułości sensora, staje się fundamentalną decyzją o charakterze wartościującym. Odpowiada człowiek. Zawsze człowiek.

## Od statycznej sygnatury do politycznej ekonomii percepcji

W świecie zaawansowanych sensorów niezwykle łatwo jest ulec pokusie mierzenia tego, co mierzalne, zamiast tego, co rzeczywiście decydujące. Można bez trudu zachwycić się chirurgiczną precyzją klasyfikacji modelu na sterylnym benchmarku, ale znacznie trudniej ocenić, czy operator działający w warunkach skrajnego stresu rozumie, kiedy maszyna ma rację, a kiedy jedynie sprawia wrażenie pewnej siebie. Algorytm może legitymować się imponującą statystyką trafień, a jednocześnie strategicznie rujnować proces decyzyjny, jeśli produkuje niebezpieczne złudzenie niezawodności. Komputer, jak każdy współczesny biurokrata, bywa szczególnie groźny wtedy, gdy myli pewność z kompetencją. To tutaj objawia się automation bias, czyli tendencja do bezkrytycznego ufania wynikom algorytmu, co prowadzi do fikcyjnej delegacji odpowiedzialności na system, którego człowiek przestał w pełni rozumieć.

Z tej perspektywy klasyfikacja sygnałów przestaje być wąsko technicznym zagadnieniem rozpoznawania wzorców, a staje się rdzeniem nowej ekonomii politycznej percepcji. W teatrze współczesnych działań informacja jest walutą, a szybkość jej obróbki wyznacza bogactwo dostępnych opcji strategicznych. Kto klasyfikuje lepiej, ten widzi wcześniej, zyskując luksus czasu, którego brakuje przeciwnikowi. Kto widzi wcześniej, ten taniej przemieszcza zasoby, skuteczniej ukrywa własne zamiary i szybciej demaskuje cudze fortele. Jest to gra o marże, w której sensor staje się najważniejszym aktywem w portfelu dowódcy. Percepcja nie jest tu neutralną obserwacją, lecz przewagą konkurencyjną, którą można wymienić na przetrwanie.

W klasyfikacji akustycznej spektrogram staje się obrazem, a sieci CNN przestają być architekturą do rozpoznawania zdjęć kotów, zaczynając pełnić rolę filtra przemocy, alarmu i atrybucji źródła. Podobną ewolucję obserwujemy w radiolokacji, gdzie następuje przejście od analizy pojedynczych

impulsów ku sekwencjom o znacznie wyższym poziomie abstrakcji. Książka wskazuje, że w świecie zwinnych radarów analiza na poziomie samych „liter" sygnałowych okazuje się dziś niewystarczająca. Dopiero sylaby, słowa i gramatyka emisji ujawniają głęboką logikę działania obcego systemu. To jeden z najciekawszych wątków współczesnego ELINT, czyli wywiadu elektronicznego opartego na przechwytywaniu sygnałów. Emiter nie jest już tylko źródłem suchych parametrów, lecz staje się nadawcą unikalnego stylu operacyjnego.

Nie odczytujemy już jedynie częstotliwości, szerokości impulsu czy czasu nadejścia, jakbyśmy wypełniali rubryki w starym rejestrze. Odczytujemy zachowanie, co sprawia, że współczesny wywiad elektroniczny coraz mniej przypomina katalog motyli, a coraz bardziej hermeneutykę maszyn. Łańcuchy Markowa i sieci LSTM nie tyle rozpoznają urządzenie, ile rekonstruują jego zwyczaj i ukrytą intencję. A zwyczaj, jak wiadomo, zdradza o naturze aktora znacznie więcej niż jakakolwiek oficjalna deklaracja. Ten zwrot od parametrów do zachowania ma kolosalne znaczenie strategiczne w dobie walki radioelektronicznej. Tradycyjne bazy sygnatur opierały się bowiem na założeniu względnej stabilności trybów pracy, co w obecnych warunkach jest anachronizmem.

Współczesny radar zwinny jest systemem, który za wszelką cenę nie chce być rozpoznany po jednej pozie. Zmienia konfiguracje, adaptuje fale i rekonfiguruje cykle pracy, przestawiając priorytety stosownie do dynamicznej sytuacji na polu walki. Próba ujęcia takiego obiektu wyłącznie przez statyczny słownik jest analogiczna do próby scharakteryzowania człowieka na podstawie jednego odruchu źrenicy. Da się to oczywiście oprawić w elegancki raport, ale trudno nazwać to rzeczywistym rozumieniem zagrożenia. Paradygmat naukowy przesuwa się zatem od ontologii sygnału ku pragmatyce emisji. Nie pytamy już tylko, jaki parametr ma impuls, lecz jaką funkcję pełni on w sekwencji i w jakim reżimie zarządzania zasobami powstał.

Właśnie dlatego open set recognition, czyli zdolność klasyfikatora do rozpoznawania zjawisk spoza znanego katalogu, staje się tak fundamentalna. Klasyfikator, który każdą nowość na siłę przypisuje do znanego wzorca, jest nie tyle inteligentny, ile niebezpiecznie zarozumiały. W wywiadzie elektronicznym pycha algorytmu to najkrótsza droga do fałszywej pewności, która na polu walki jest po prostu kosztowną formą ignorancji. Meaningful human control, czyli standard realnego nadzoru nad systemem, wymaga, by maszyna potrafiła komunikować własną niepewność. Bez tego człowiek w pętli decyzyjnej staje się jedynie ceremonialnym dodatkiem do procesów, których nie kontroluje. Prawdziwa inteligencja systemu objawia się w jego pokorze wobec danych, których nie potrafi jednoznacznie zinterpretować.

Dochodzimy tu do problemu planowania misji i zarządzania zasobami, czyli dynamicznej administracji czasem i mocą sensora w warunkach wewnętrznej konkurencji zadań. W klasycznych wyobrażeniach dowódca wydaje rozkaz, radar pracuje, a informacja płynie szerokim strumieniem

do sztabu. W realnym systemie wielofunkcyjnym nic nie jest jednak tak proste ani tak oczywiste. Każda milisekunda czasu pracy wiązki, każdy margines mocy i każda decyzja o śledzeniu zamiast przeszukiwania jest wyborem ekonomicznym dokonywanym pod presją niedoboru. Radar stał się przedsiębiorstwem działającym w warunkach permanentnego kryzysu alokacyjnego. Nie może zrobić wszystkiego naraz, więc musi handlować własną uwagą, która jest jego najcenniejszym zasobem.

Kategoria Quality of Service (QoS) ma w tym obszarze znaczenie o wiele głębsze niż sugerowałby to suchy, techniczny żargon. QoS jest w istocie próbą przetłumaczenia niewspółmiernych jakości na wspólną skalę użyteczności operacyjnej. Błąd śledzenia, zasięg wykrycia, opóźnienie odświeżenia i pewność klasyfikacji nie dają się bezpośrednio porównać bez wspólnego mianownika. Tradycyjne podejście zawodzi, toteż trzeba je przełożyć na ekonomię celu misji, co stanowi o aksjologii progu alarmowego. Gdy inżynierowie mówią o warunkach optymalności, nie opisują wyłącznie problemu matematycznego. Opisują politykę wyboru tego, co system uzna za warte swojego ograniczonego czasu. A czas w architekturach sensorowych jest dobrem równie rzadkim jak uwaga w gospodarce cyfrowej.

Nowoczesne zarządzanie zasobami w sieciach sensorów należy czytać jako teorię państwa w miniaturze. Każdy system musi rozstrzygać wewnętrzny konflikt pomiędzy funkcjami, które wszystkie zgłaszają absolutne pierwszeństwo. Śledzenie alarmuje, że bez ciągłości kontaktu stracimy obiekt, podczas gdy przeszukiwanie odpowiada, że bez pokrycia przestrzeni nie zobaczymy nowego zagrożenia. Komunikacja żąda przepustowości, walka elektroniczna chce dominacji w spektrum, a identyfikacja domaga się czasu i danych. Każda z tych funkcji wygłasza racjonalną mowę końcową, a procesor, niczym znużony minister finansów, musi ostatecznie ogłosić budżet. W tym miejscu algorytmy typu Q-RAM nie są jedynie technikami obliczeniowymi, lecz konstytucjami maszynowej suwerenności.

Ustalają one, jak przekładać dobro całości na konkretną dystrybucję środków w warunkach konfliktu interesów. Z perspektywy prawnej i politycznej jest to fascynujące, ponieważ pokazuje, że nawet tam, gdzie rządzi algebra, ostatecznie powraca kwestia wartości. Nie istnieje coś takiego jak neutralna alokacja w warunkach niedoboru; każda architektura optymalizacyjna jest zakamuflowaną teorią priorytetów. Nawet jeśli projektant woli mówić o parametrach technicznych, to system i tak zdradza swoim działaniem, co kocha bardziej. Systemy, podobnie jak imperia, najwięcej mówią o sobie nie w oficjalnych deklaracjach, lecz w strukturze swojego budżetu. To właśnie tam ukryta jest prawdziwa hierarchia ważności, która steruje okiem maszyny.

Szczególnej uwagi wymaga nurt Track before Detect (TbD), ponieważ doskonale ilustruje on zmianę paradygmatu od prostego wykrywania do kumulacji słabej wiedzy. Konwencjonalna logika

mówiła, że najpierw należy przekroczyć próg detekcji, a dopiero potem można zacząć śledzenie obiektu. TbD odwraca tę kolejność, pozwalając gromadzić niepewne, słabe i zaszumione echa na przestrzeni kolejnych obserwacji. Dzięki temu obiekt wyłania się nie z jednego triumfalnego sygnału, lecz z cierpliwie budowanej struktury prawdopodobieństwa. Epistemologicznie jest to ruch bardzo doniosły, oznaczający przejście od kultury punktowego dowodu ku kulturze akumulacji subtelnych przesłanek. W świecie dronów i niskiego stosunku sygnału do szumu to nie fanaberia, lecz warunek dostrzeżenia tego, co niewidoczne.

Oczywiście koszt takiego podejścia jest wysoki, gdyż TbD bywa obliczeniowo ciężkie i trudne do wdrożenia w czasie rzeczywistym. Lecz właśnie ta kosztowność pokazuje istotę współczesnego pola walki informacyjnej, gdzie widzenie stało się bezpośrednią funkcją mocy obliczeniowej. Moc ta z kolei zależy od instytucjonalnej zdolności do inwestowania w zarządzanie niepewnością. Kto nie chce płacić za niepewność, ten później płaci znacznie wyższą cenę za własną ślepotę. Warto w tym miejscu przywołać intuicję z książki, która dla całego dalszego wywodu okaże się fundamentalna. Redukcja złożoności, czyli proces upraszczania rzeczywistości przez AI w celu umożliwienia reakcji operacyjnej, musi być prowadzona świadomie, by nie stała się jedynie narzędziem do produkowania wygodnych kłamstw.

## Filologia wojenna: Od analizy sygnału do logiki zachowania

Nadmiar danych nie daje automatycznie prawdy; sterta cegieł nie staje się domem bez projektu. Informacja w systemach maszynowych posiada ambiwalentną naturę, potrafiąc jednocześnie alarmować, porządkować chaos, ale i skutecznie usypiać czujność operatora. Kiedy implementujemy AI w sensorach ELINT czy systemach rozpoznania nuklearnego, nie powinniśmy ulegać fetyszyzacji tego procesu jako cudownego rozszerzenia ludzkiego poznania. W rzeczywistości mamy do czynienia z redukcją złożoności, czyli procesem brutalnego upraszczania rzeczywistości przez algorytmy w celu umożliwienia szybkiego działania operacyjnego. Ta techniczna wygoda ma jednak swoją cenę, ponieważ każda taka operacja odbywa się kosztem selektywnego pomijania tych fragmentów świata, które nie pasują do przyjętego modelu.

Etyka systemów autonomicznych nie zaczyna się od naiwnych pytań, czy maszyna potrafi nas zrozumieć lub poczuć empatię. Prawdziwe wyzwania moralne rodzą się z twardszej analizy tego, co dany system czyni niewidzialnym i komu w ostatecznym rozrachunku przekazuje strategiczną przewagę. Musimy pytać, kogo algorytm narazi na fałszywy alarm, czyją pomyłkę uzna za akceptowalny koszt i czyje życie zostanie beznamiętnie wpisane do kolumny szkód ubocznych.

Właśnie w tym punkcie kończy się infantylna opowieść o neutralnej technologii, a zaczyna poważna rozmowa o fundamentach naszej cywilizacji. Dopiero po odrzuceniu tych złudzeń możemy przejść do szczegółowej analizy architektury problemu, obejmującej obliczenia kwantowe oraz próby algorytmizacji prawa wojennego.

Klasyfikacja sygnałów, stanowiąca zmysł nowoczesnego systemu ELINT, staje się dziś filologią wojenną, wymagającą znacznie więcej niż tylko biernego nasłuchu. Współczesnemu analitykowi nie wystarcza już samo zarejestrowanie impulsu, musi on bowiem rozpoznać idiom emitera, jego specyficzne nawyki oraz unikalną składnię operacyjną. Przesuwamy środek ciężkości z analizy pojedynczego parametru ku interpretacji całych sekwencji, co oznacza fundamentalne przejście od atomu sygnału ku logice zachowania maszyny. To ruch o znaczeniu nie tylko technicznym, lecz przede wszystkim epistemologicznym, zmieniający sposób, w jaki konstruujemy wiedzę o przeciwniku. Tradycyjny wywiad elektroniczny przypominał dawnego bibliotekarza, który święcie wierzył, że każdą książkę rozpozna po grubości papieru, kolorze okładki czy kroju czcionki.

W świecie zwinnych radarów ten bibliotekarz przegrywa, gdyż dzisiejszy emiter potrafi rekonfigurować parametry w ułamkach sekund. Urządzenie pozostaje rozpoznawalne dopiero wtedy, gdy zaczynamy czytać jego praktykę jako strukturę uporczywie powracających wyborów, a nie zbiór statycznych cech fizycznych. Tak właśnie należy rozumieć hierarchiczny model emisji, w którym pojęcia takie jak litery, sylaby czy komendy nie są jedynie metaforą, lecz próbą uchwycenia wielopoziomowej organizacji działania. Współczesny paradygmat naukowy odchodzi więc od prostego katalogowania sygnatur na rzecz traktowania sygnału jako złożonego języka zachowań maszynowych. Ta zmiana znajduje swoje odbicie w szerokiej debacie o interpretowalności systemów AI, gdzie standardy techniczne zaczynają przenikać się z teorią komunikacji.

NIST w najnowszych pracach akcentuje konieczność stałego monitorowania systemów AI już po ich wdrożeniu w realne warunki. Dzieje się tak dlatego, że zachowanie systemu w świecie nigdy nie kończy się na właściwościach modelu opisanych w sterylnych warunkach laboratoryjnych. Musimy brać pod uwagę automation bias, czyli tendencję operatorów do bezkrytycznego ufania wynikom generowanym przez algorytmy, nawet przy ich ewidentnej błędności. Bez ciągłego nadzoru ryzykujemy, że meaningful human control, czyli standard nadzoru gwarantujący realną możliwość interwencji człowieka, stanie się jedynie pustym hasłem. Prawdziwa kontrola wymaga bowiem zrozumienia, że system w interakcji z otoczeniem ewoluuje, tworząc nowe, nieprzewidziane wcześniej wzorce aktywności.

Przewaga modeli sekwencyjnych, takich jak sieci LSTM, nad naiwnym rozpoznawaniem słownikowym nie wynika z jakiejś magicznej głębi sieci neuronowych. Bierze się ona z prostego faktu, że nowoczesny radar nie emituje jednego, niezmiennego numeru rejestracyjnego, lecz

prowadzi ciągłą politykę sygnałową. Polityka ta posiada własny rytm, wewnętrzną hierarchię oraz ściśle określoną funkcję taktyczną, której nie da się uchwycić za pomocą prostych tabel. Analiza na najniższym poziomie, ograniczona do samych impulsów, daje wyniki bliskie losowości, ponieważ różne systemy często współdzielą te same elementy bazowe. Dopiero na poziomie wyższej gramatyki, tam gdzie impulsy układają się w logiczne ciągi, zaczyna ujawniać się rzeczywista, niepodrabialna odrębność konkretnego emitera.

To spostrzeżenie ma ogromne znaczenie praktyczne, definiując na nowo pojęcie przewagi w nowoczesnym wywiadzie elektronicznym. Oznacza ono, że sukces coraz rzadziej będzie wynikał z samego faktu gromadzenia gigantycznych zbiorów sygnałów, a coraz częściej z jakości modeli zdolnych do uchwycenia ich wewnętrznego uporządkowania. W dzisiejszej gospodarce danych liczba rekordów stała się towarem tanim i powszechnym, wręcz trywialnym w swojej masowości. Prawdziwa wartość kryje się w zdolności do interpretacji, w owej hermeneutyce maszyn, która pozwala odróżnić istotny sygnał od szumu tła. Kto nie chce płacić za niepewność płynącą z konieczności ciągłej interpretacji, ten prędzej czy później zapłaci znacznie wyższą cenę za swoją strategiczną ślepotę.

## Epistemiczna pokora: Dlaczego AI musi umieć powiedzieć nie wiem

Współczesny wywiad elektroniczny coraz mniej przypomina katalog motyli, a coraz bardziej hermeneutykę maszyn, gdzie kluczowa staje się umiejętność rozróżniania subtelnych stylów działania. Można by rzec, z właściwą tej branży odrobiną wisielczego humoru, że nowoczesny radar nie zdradza się już tym, jak wygląda, lecz tym, jak chodzi. Kto nie opanuje sztuki czytania tego cyfrowego chodu, ten w warunkach walki elektronicznej zostaje z bardzo drogim zbiorem statycznych, nic nieznaczących szczegółów. W świecie rzeczywistym system nie może bowiem zakładać, że każdy napotkany obiekt musi koniecznie należeć do jednej ze znanych mu wcześniej szufladek. Klasyfikator zamknięty jest rozwiązaniem niezwykle wygodnym dla twórców akademickich benchmarków, lecz bywa fatalny w skutkach dla operacyjnej rozwagi na polu walki.

Jeśli każdy nieznany sygnał zostaje na siłę przypisany do którejś z wyuczonych klas, model wcale nie zwiększa naszego bezpieczeństwa, lecz produkuje wyjątkowo wyrafinowaną pomyłkę. W obszarach wysokiego ryzyka błąd ten jest szczególnie groźny, ponieważ przybiera postać uporządkowaną, policzoną i nienagannie sformatowaną. Trzeba przyznać, że my, ludzie, mamy niebezpieczną słabość do elegancko sformatowanych błędów, które mylimy z obiektywną prawdą. Open set recognition, czyli praktyka polegająca na identyfikowaniu, czy sygnał należy do znanej

klasy, czy też stanowi nieznany wcześniej byt, jest zatem nie tylko techniką, ale wręcz instytucją epistemicznej pokory. Nakazuje ona systemowi prostą, lecz fundamentalną zasadę: nie udawaj wiedzy tam, gdzie jej po prostu nie posiadasz.

Ta reguła pozostaje głęboko zbieżna z wytycznymi dla wyjaśnialnej sztucznej inteligencji sformułowanymi przez NIST, a konkretnie z zasadą knowledge limits. Zgodnie z nią system powinien operować wyłącznie w warunkach, dla których został zaprojektowany, lub w momentach, gdy osiąga wystarczającą pewność odpowiedzi. W gruncie rzeczy mamy tutaj do czynienia z bardzo staroświecką, niemal zapomnianą cnotą intelektualną. Dobra maszyna, podobnie jak dobry ekspert, powinna przede wszystkim umieć powiedzieć „nie wiem" bez poczucia porażki. W epoce technologicznej pychy, gdzie algorytmy mają rzekomo znać odpowiedzi na każde pytanie, jest to postawa niemal kontrkulturowa.

Na tym tle koncepcja Track before Detect, w skrócie TbD, jawi się jako logiczne rozwinięcie tej samej, ostrożnej filozofii poznawczej. Nie chodzi w niej o spektakularne wykrycie obiektu, który z impetem przekracza ustalony próg detekcji i domaga się natychmiastowej uwagi. TbD, czyli praktyka polegająca na cierpliwym wydobywaniu sensu z danych zbyt słabych, by samodzielnie uprawomocnić wniosek, to epistemologia głęboko nieufna wobec ostentacyjnej pewności. Zakłada ona, że obiekt może być realny i groźny, nawet jeśli nie pojawia się w postaci jednej, jasnej i klarownej deklaracji systemu. Wymaga to całkowania słabych śladów w czasie i spinania ich w złożoną strukturę prawdopodobieństwa.

Dopiero taka akumulacja subtelnych sygnałów pozwala, by obecność celu wyłoniła się z szumu w sposób wiarygodny i przemyślany. Jest to rozwiązanie szczególnie cenne w pasywnej radiolokacji oraz w wykrywaniu małych dronów, gdzie niski stosunek sygnału do szumu czyni klasyczne progi detekcyjne całkowicie zawodnymi. W tym sensie TbD stanowi metodologiczny bunt przeciwko tyranii pierwszego wrażenia, która tak często zwodzi operatorów systemów obronnych. Algorytm nie pyta już naiwnie, co widać w tej konkretnej milisekundzie, lecz docieka, jaki wzorzec zaczyna się konstytuować w dłuższym przedziale czasu. Dla profesjonalnego czytelnika warto podkreślić, że nie jest to wyłącznie drobny detal natury technicznej.

Mamy tu do czynienia z przesunięciem całego modelu rozumienia zjawisk fizycznych przez systemy informatyczne. Zamiast logiki natychmiastowego, często błędnego rozpoznania, wdrażamy logikę cierpliwej kumulacji informacji i ważenia dowodów. Zamiast kultu jednorazowego, twardego dowodu, wprowadzamy politykę probabilistycznego dojrzewania wniosku do momentu jego pełnej wiarygodności. W praktyce systemów bezpieczeństwa jest to istotny postęp cywilizacyjny, choć oczywiście okupiony wymiernymi kosztami obliczeniowymi oraz opóźnieniami. Tu jednak należy

zachować zdrowy sceptycyzm, ponieważ TbD nie jest magicznym zaklęciem, które obala twarde ograniczenia fizyki.

Każda metoda tego typu niesie ze sobą potężny ciężar obliczeniowy, problemy z przetwarzaniem w czasie rzeczywistym oraz trudności przy dużym zagęszczeniu śladów. Jest to niezwykle ważne, gdyż współczesne narracje technologiczne lubią opowiadać o nowych algorytmach tak, jakby były one darmowym dodatkiem do rzeczywistości. Tymczasem w inżynierii nie istnieje widzenie bez kosztu, a każda próba zajrzenia głębiej w szum ma swoją cenę. Gdy system gromadzi słabe echa przez wiele cykli, płaci za to pamięcią, mocą obliczeniową i ryzykiem błędnej konsolidacji danych. W ekonomii nazwalibyśmy to zjawisko gwałtownym wzrostem kosztu krańcowego wiarygodności.

Im bardziej desperacko chcemy wyłuskać informację z chaosu, tym więcej musimy zapłacić za każdą dodatkową jednostkę pewności. Właśnie dlatego TbD znakomicie ilustruje szerszą tezę, że nowoczesna AI w systemach sensorowych nie jest fabryką wszechwiedzy. Jest ona raczej kosztowną sztuką kupowania nieco lepszego rozeznania w warunkach nieusuwalnej i wszechobecnej niepewności. Kto obiecuje nam więcej, ten handluje już nie nauką, lecz niebezpiecznym techno-mesjanizmem, który mami wizją nieomylności. A techno-mesjanizm ma tę przykrą cechę, że zwykle krzyczy najgłośniej tam, gdzie najmniej rozumie realną cenę własnych cudów.

Jeszcze wyraźniej widać to w skomplikowanym obszarze rozpoznania nuklearnego, gdzie margines błędu praktycznie nie istnieje. Zestawienie stacjonarnych systemów wykrywania zagrożeń radiologicznych z powietrznymi platformami bezzałogowymi pokazuje, jak daleko posunęła się fuzja danych. W rozwiązaniach stacjonarnych sam detektor gamma często nie wystarcza, gdyż mierzy on jedynie chaotyczną mieszaninę tła i potencjalnych źródeł. Dopiero sprzężenie tych odczytów z trajektoriami ruchu osób, pozyskanymi z kamer, otwiera możliwość przypisania promieniowania konkretnej jednostce w tłumie. Wariant lotniczy z kolei minimalizuje narażenie ludzi, ale natychmiast odsłania własne, dotkliwe ograniczenia pomiarowe.

Drony zmagają się z nasyceniem detektorów, zanikiem słabych sygnałów na wysokości oraz opóźnieniami wynikającymi z konieczności transmisji danych. To bardzo pouczający przypadek, który pokazuje, że w systemach krytycznych nie istnieje coś takiego jak czyste, nieskażone źródło wiedzy. Wiedza operacyjna jest zawsze wypadkową wielu kanałów, modeli korekcyjnych oraz nieustannej kompensacji błędów. Fuzja danych, czyli proces integrowania wielu źródeł w celu uzyskania spójniejszego obrazu sytuacji, nie jest luksusem, lecz warunkiem sensownego działania. Jednocześnie jest to proces politycznie i prawnie brzemienny w skutkach dla całego społeczeństwa.

Każdy dodatkowy sensor wpięty w sieć oznacza przecież nowy punkt odpowiedzialności oraz nowy wektor potencjalnego błędu. Właśnie w tym miejscu etyka sztucznej inteligencji spotyka się z

prawem ochrony prywatności oraz zasadą proporcjonalności działań państwa. Komisja Europejska, opisując wejście w życie AI Act, wyraźnie podkreśliła, że regulacje mają godzić rozwój technologii z ochroną bezpieczeństwa i praw podstawowych. W przeciwnym razie modne hasło human in the loop staje się jedynie liturgią odpowiedzialności bez odpowiedzialności, gdzie człowiek podpisuje decyzje, których nie jest w stanie pojąć.

## Polityczna ekonomia sensora: dlaczego algorytmy nie zastąpią sumienia

W obszarze systemów wysokiego ryzyka napięcie między precyzją a błędem nie jest jedynie technologicznym naddatkiem, lecz samą istotą technologii odczytaną w języku instytucji. Warto zauważyć, że rozpoznanie nuklearne stanowi szczególne laboratorium odpowiedzialności, w którym wyjątkowo wyraźnie rozpada się infantylne marzenie o całkowicie neutralnym sensorze. Sam wybór progu alarmowego, czyli momentu, w którym system uznaje sygnał za realne zagrożenie, jest w istocie fundamentalną decyzją aksjologiczną. Aksjologia progu alarmowego, czyli uznanie, że wybór czułości sensora jest decyzją wartościującą, a nie tylko techniczną, uświadamia nam, że każdy system detekcji opiera się na kruchym kompromisie. Im niżej zawiesimy tę poprzeczkę, tym większa staje się szansa na wykrycie rzeczywistego niebezpieczeństwa, ale jednocześnie gwałtownie rośnie liczba fałszywych alarmów oraz kosztów operacyjnych. Pociąga to za sobą nie tylko paraliż decyzyjny, lecz także potencjalne naruszenia godności osób niesłusznie branych na celownik algorytmu. Z kolei podniesienie progu, choć pozornie porządkuje świat i redukuje szum, drastycznie zwiększa ryzyko przeoczenia obiektu, który może zakończyć historię znanego nam świata.

W tym klinczu nie istnieje żadne czysto techniczne wyjście poza konflikt wartości, a jedyną uczciwą drogą pozostaje świadome projektowanie kompromisu. Dlatego należy przyznać rację tym, którzy traktują sztuczną inteligencję i algorytmy wyłącznie jako wsparcie dla ludzkiego sądu, a nie jego ostateczne zastąpienie. Człowiek pozostaje w tym układzie niezbędny nie po to, by z miną ceremonialnego patriarchy wpatrywać się w migoczący ekran, lecz by instytucjonalnie autoryzować granice akceptowalnego ryzyka. Maszyna bez trudu policzy rozkład prawdopodobieństwa, lecz nigdy nie rozstrzygnie samodzielnie, ilu fałszywie zatrzymanych obywateli jesteśmy gotowi uznać za dopuszczalny koszt ochrony infrastruktury krytycznej. Taka decyzja należy do porządku politycznego, a nie do architektury obliczeniowej. Jeśli system udaje, że jest inaczej, mamy do czynienia nie z postępem, lecz z niebezpiecznym outsourcingiem sumienia. W tym miejscu human in the loop staje się liturgią odpowiedzialności bez odpowiedzialności.

Współczesna technologia nie jest zjawiskiem przyrodniczym, które spada na nas niespodziewanie niczym deszcz, lecz stanowi bezpośredni skutek konkretnych inwestycji, bodźców prawnych i form organizacji. Traktowanie rozwoju AI jako metafizycznej konieczności jest błędem poznawczym, ponieważ każda automatyzacja wynika z określonych decyzji politycznych oraz fiskalnych. Ten pogląd należy bezwzględnie rozciągnąć na sferę obronności, wywiadu elektronicznego ELINT oraz skomplikowanych architektur sensorowych. Współczesny wywiad elektroniczny coraz mniej przypomina katalog motyli, a coraz bardziej hermeneutykę maszyn. Także tutaj nie rządzi żadne losowe przeznaczenie ani nieubłagane fatum, przed którym musimy skłonić głowę. To, czy dana instytucja zainwestuje w wyjaśnialną sztuczną inteligencję, monitoring po wdrożeniu czy procedury audytowe, jest wyborem ustrojowym o głębokich konsekwencjach. Wyborem jest również to, czy mierzymy realną wartość operacyjną systemu, czy jedynie ulegamy współczesnemu teatrowi wskaźników KPI, który karmi nasze poczucie bezpieczeństwa pustymi liczbami.

Koncepcja Dobrego Państwa słusznie zwraca uwagę, że systemy SI należy traktować jako infrastrukturę krytyczną, mierząc ich skuteczność poprzez outcome KPIs, czyli wskaźniki realnych efektów, a nie wyłącznie parametry robocze. Każdy projekt technologiczny musi być osadzony w szerszym szkielecie strategicznym, co z pozoru brzmi jak uwaga biznesowa, lecz w istocie jest stwierdzeniem głęboko politycznym. W systemach wysokiego ryzyka wskaźnik techniczny, który nie przekłada się na realne bezpieczeństwo, staje się formą groźnego samooszustwa organizacyjnego. Można przecież posiadać estetyczny interfejs i jednocześnie operować w ramach skrajnie wadliwej architektury decyzji. A marketing, jak wiadomo, potrafi sprzedać nawet mgłę jako infrastrukturę krytyczną, byle tylko nadać jej dashboard. Historia nowoczesności aż za dobrze zna ten typ triumfu formy nad treścią. To prowadzi nas bezpośrednio do zagadnienia planowania misji, gdzie algorytmy genetyczne służą optymalizacji rozmieszczenia sensorów i minimalizacji błędów estymacji.

Pod powierzchnią tych technicznych operacji działa jednak znacznie głębsza logika, ponieważ planowanie misji to moment, w którym epistemologia spotyka się z geometrią. Nie wystarczy bowiem dysponować dobrym modelem klasyfikacji, jeśli nie potrafimy tak ulokować sensorów, aby powstała realna przewaga informacyjna w czasie i przestrzeni. Jest to jedna z najbardziej niedocenianych prawd o sztucznej inteligencji: modele nie działają w próżni, a ich jakość zależy od materialnej topologii świata. Przeszkody terenowe, linie widzenia, propagacja sygnału czy budżet energetyczny to twarde ograniczenia, których nie przeskoczy żadna linia kodu. Każda opowieść o wszechmocy AI, która pomija fizyczną infrastrukturę i wąskie gardła, takie jak dostęp do półprzewodników czy chmury obliczeniowej, jest z gruntu fantazją. Inteligencja obliczeniowa jest bowiem tylko tak dobra, jak materialna architektura jej działania. Nawet najbardziej błyskotliwy

algorytm bywa bezradny, jeśli sieć sensorów została rozstawiona jak meble w mieszkaniu urządzanym po ciemku.

Z tego wynika ścisły związek między planowaniem misji a zarządzaniem zasobami (Resource Management), czyli dynamiczną administracją czasem i mocą obliczeniową sensora w warunkach wewnętrznej konkurencji zadań. Zarządzanie tymi zasobami przestało być jedynie problemem inżynierii wydajności, stając się w istocie problemem konstytucji uwagi systemu. Maszyna musi stale rozstrzygać, co w danej chwili zasługuje na czas antenowy, moc obliczeniową oraz przepustowość łącza. W tym miejscu kategoria jakości usług (Quality of Service – QoS), czyli mechanizm zapewniania jakości usług poprzez mapowanie różnych jakości na wspólną skalę użyteczności, staje się pojęciem filozoficznie płodnym. System dokonuje tego, co państwa i korporacje robią od dawna w innych dziedzinach życia społecznego. Przekłada on niewspółmierne dobra na jeden porządek porównywalności, tworząc specyficzną polityczną ekonomię sensora, w której żadna alokacja nie jest etycznie niewinna.

Wybór konkretnej funkcji użyteczności przesądza o tym, czy system będzie bardziej konserwatywny, agresywny, czy może gotowy do ryzykownej ekspansji pola obserwacji. To już nie jest matematyka uprawiana dla samej matematyki, lecz matematyka jako język preferencji instytucji, z którego człowiek nie może zostać arbitralnie wycięty. Ktoś musi przecież wziąć odpowiedzialność za to, jaki świat wartości wpisano do funkcji celu, która steruje zachowaniem automatu. Szczególnie ciekawa jest tutaj analogia do ekonomii dobrobytu i warunków optymalności Karusha-Kuhna-Tuckera (KKT), które dla laika brzmią jak neutralna konieczność obliczeniowa. W praktyce są one jednak formalnym opisem punktu, w którym dalsze przesuwanie zasobów nie zwiększa już łącznej użyteczności całego systemu. Brzmi to racjonalnie i elegancko, dopóki nie zaczniemy pytać, o czyją użyteczność właściwie chodzi, w jakim horyzoncie czasowym jest ona mierzona i z jaką wagą traktuje się ryzyko.

Każda odpowiedź na tak postawione pytania oznacza przyjęcie ukrytej normatywności, co stanowi jedną z najważniejszych lekcji płynących z analizy systemów autonomicznych. Nawet tam, gdzie pozornie rządzą wyłącznie bezduszne algorytmy alokacyjne, nieuchronnie powraca fundamentalne pytanie etyczne. Nie dzieje się tak dlatego, że matematyka sama w sobie jest niemoralna, lecz dlatego, że zawsze optymalizuje ona coś, co zostało wcześniej zdefiniowane przez człowieka. System może przecież z matematyczną perfekcją rozwiązać całkowicie błędnie postawiony problem. A perfekcyjne rozwiązywanie źle postawionych problemów jest, trzeba przyznać, jedną z najstarszych specjalności zarówno biurokracji, jak i dobrze opłacanej konsultacji strategicznej. Komputer, jak każdy współczesny biurokrata, bywa szczególnie groźny wtedy, gdy myli pewność obliczeń z

kompetencją merytoryczną. Maszyna robi to po prostu szybciej, sprawiając, że nasze błędy stają się natychmiastowe i nieodwołalne. Odpowiada człowiek. Zawsze człowiek.

## Kwantowa inżynieria: Jak fizyka ratuje separację celów

Ryzykowne intelektualnie, a zarazem najbardziej płodne fragmenty współczesnej myśli technicznej zaczynają się tam, gdzie inżynieria przestaje zadowalać się własnym, hermetycznym językiem i zaczyna bezczelnie pożyczać pojęcia z fizyki teoretycznej. Zasada wykluczenia Pauliego, kwantowa nierozróżnialność, obliczenia adiabatyczne czy model Isinga to cały arsenał pojęć, który laikowi może trącić albo metafizyką laboratorium, albo egzaltacją konferencyjnej prezentacji. W rzeczywistości jednak te zapożyczenia odpowiadają na bardzo trzeźwy, niemal brutalny problem: jak nie zgubić odrębności poszczególnych obiektów w świecie gęstym, zaszumionym i skrajnie dynamicznym. Chodzi o to, by rozdzielić ślady tam, gdzie klasyczne algorytmy mają naturalną skłonność do ich fatalnego zlewania, co w praktyce oznacza utratę kontroli nad sytuacją. Musimy bowiem ujarzmić eksplozję kombinatoryczną w momentach, gdy liczba możliwych przypisań danych do obiektów zaczyna rosnąć znacznie szybciej niż cierpliwość procesora. W przeciwnym razie system po prostu przestaje nadążać za rzeczywistością, a my zostajemy z cyfrowym chaosem zamiast precyzyjnego obrazu sytuacji.

Współczesny paradygmat naukowy w dziedzinie sztucznej inteligencji dla sensorów nie rozwija się wcale linearnie, od prostszych modeli ku coraz większej liczbie warstw sieci neuronowych. Rozwija się on raczej przez systematyczną kradzież pojęć z innych, często odległych dziedzin, co jest zresztą oznaką intelektualnej dojrzałości, a nie słabości. Kradnie się z rachunku prawdopodobieństwa, teorii decyzji, ekonomii dobrobytu, a nawet z prawa i filozofii moralnej, byle tylko znaleźć skuteczny opis świata. Niedojrzała dyscyplina naiwnie wierzy, że wystarczy jej własny, branżowy słownik, podczas gdy dojrzała przyznaje, że rzeczywistość bywa bardziej złośliwa niż sztywne granice wydziałów uniwersyteckich. I słusznie, bo cel przecinający się z innym celem na ekranie radaru nie pyta przecież, czy laboratorium analizujące ten ruch należy do instytutu fizyki, czy informatyki. A marketing, jak wiadomo, potrafi sprzedać nawet mgłę jako infrastrukturę krytyczną, byle tylko nadać jej efektowny dashboard i odpowiednio modną nazwę.

Kwantowa inspiracja w systemach śledzenia nie pełni roli estetycznego ozdobnika, gdyż problem monitorowania wielu celów zawiera w sobie własną, głęboką ontologię nierozróżnialności. Gdy sensory dostarczają nam surowych danych kinematycznych bez twardych identyfikatorów, system traci prostą pewność, które echo należy do konkretnego, fizycznego obiektu. Właśnie tam rodzi się zjawisko koalescencji śladów, czyli sytuacja, w której dwa blisko położone obiekty zaczynają w

reprezentacji danych wyglądać jak jeden, niebezpiecznie stopiony byt. Klasyczne filtry i mechanizmy asocjacji mogą wtedy zachowywać się jak spanikowany urzędnik w płonącym archiwum, który w pośpiechu miesza teczki różnych petentów. Komputer, jak każdy współczesny biurokrata, bywa szczególnie groźny wtedy, gdy myli pewność z kompetencją, produkując eleganckie, lecz całkowicie błędne wyniki. Każde późniejsze próby uporządkowania takiego bałaganu są już tylko rozpaczliwym ratowaniem twarzy po bezpowrotnej utracie faktów.

Adaptacja zasady wykluczenia Pauliego do problematyki śledzenia wielu celów jest zatem ruchem niezwykle inteligentnym, choć na pierwszy rzut oka może wydawać się zbyt abstrakcyjny. Nie chodzi tu oczywiście o to, że śledzony samolot staje się nagle fermionem, lecz o wykorzystanie formalnej antysymetrii do wymuszenia pewnych logicznych ograniczeń w przestrzeni stanów. Wprowadzamy do modelu matematyczną barierę, która odpowiada naszej podstawowej fizycznej intuicji, przy czym kluczowe jest tu zachowanie odrębności bytów. Wszak dwa rozciągłe obiekty nie powinny zajmować tego samego miejsca w tym samym czasie, co w klasycznych modelach bywa nagminnie ignorowane. Jeśli model probabilistyczny dopuszcza ich zlanie się w jeden punkt, to nie jest on neutralnym narzędziem, lecz konstrukcją wadliwą względem samej natury problemu. Wprowadzenie efektu podobnego do pauliowskiego wcięcia działa więc jak twardy bezpiecznik przeciwko fikcyjnej identyczności obiektów, które w rzeczywistości pozostają osobne.

Trzeźwość osądu nakazuje jednak zachować najwyższą ścisłość i nie ulegać pokusie taniego mistycyzmu, który tak często towarzyszy wszystkiemu, co nosi przymiotnik „kwantowy”. Tego rodzaju inspiracja nie oznacza wcale, że makroskopowe śledzenie obiektów staje się nagle częścią mechaniki kwantowej, lecz wskazuje, że jej aparat formalny bywa przydatny do reprezentowania ograniczeń w procesie estymacji, czyli wyznaczania stanu obiektów. To rozróżnienie ma kluczowe znaczenie, ponieważ współczesna debata technologiczna cierpi na słabość do językowej inflacji i nadużywania wielkich słów. Wystarczy nazwać coś kwantowym, a już połowa sali zaczyna oddychać z nabożnym lękiem, jakby obcowała z jakąś wyższą, nieodgadnioną ontologią. Tymczasem prawdziwa siła tej propozycji jest znacznie bardziej surowa i sprowadza się do skutecznej redukcji błędów tam, gdzie klasyczne rozwiązania tracą rozdzielczość. Właśnie ten pragmatyczny aspekt należy podkreślać z całą mocą, jeśli tekst ma pozostać ekspercki, a nie popaść w tani, konferencyjny szamanizm. Skuteczność w scenariuszach zagęszczonych, gdzie tradycyjne algorytmy kapitulują przed natłokiem danych, jest najlepszym dowodem na sensowność takich interdyscyplinarnych poszukiwań.

# Kwantowa rewolucja w optymalizacji: od fizyki do logiki

Podobnie rzecz ma się z obliczeniami kwantowymi, które w powszechnym odbiorze dryfują niebezpiecznie blisko terytoriów zarezerwowanych dla czystej magii. W książce pojawiają się wprawdzie fundamentalne pojęcia, takie jak superpozycja, splątanie czy algorytm Grovera, niemniej wymagają one od nas surowej dyscypliny interpretacyjnej. Rejestr n kubitów opisuje przestrzeń stanów rosnącą wykładniczo wraz z ich liczbą, co rozpala wyobraźnię, lecz nie oznacza automatycznego przyspieszenia każdego zadania. To jeden z najczęstszych błędów popularyzacyjnych, bowiem kwanty nie są uniwersalnym środkiem na złożoność obliczeniową. Dają one wymierne przewagi jedynie dla ściśle określonych klas problemów, przy założeniu zastosowania odpowiednich konstrukcji algorytmicznych. Marketing, jak wiadomo, potrafi sprzedać nawet mgłę jako infrastrukturę krytyczną, byle tylko nadać jej efektowny interfejs.

W przypadku algorytmu Grovera mówimy o przyspieszeniu kwadratowym przy przeszukiwaniu nieuporządkowanych baz danych, a nie o cudownym zniesieniu wszelkich ograniczeń fizyki. Tę ostrożność warto zachować szczególnie dzisiaj, gdy amerykański NIST od lat prowadzi standaryzację kryptografii postkwantowej w obawie przed nowymi możliwościami maszyn. Nie dzieje się to jednak dlatego, że jutro każda złożoność przestanie istnieć; wszak chodzi przede wszystkim o konieczność przebudowy fundamentów bezpieczeństwa cyfrowego. Dyskusja ta dawno już opuściła bezpieczne salony futurologii, wchodząc w fazę twardej, instytucjonalnej adaptacji do nowej rzeczywistości. Pokazuje to dobitnie, że obliczenia kwantowe przestały być abstrakcją bez punktu styku z rzeczywistością, a stały się realnym czynnikiem w grze o informacyjną dominację.

Prawdziwa stawka pojawia się jednak w momencie osadzenia tych technologii w konkretnym problemie asocjacji danych. Asocjacja danych, czyli praktyka polegająca na przypisywaniu nowych sygnałów do już zidentyfikowanych obiektów, stanowi jądro współczesnych systemów śledzenia. Gdy maszyna musi rozstrzygnąć, które echa radarowe odpowiadają konkretnym śladom, przestrzeń możliwych kombinacji zaczyna gwałtownie i bezlitośnie eksplodować. Każdy praktyk pracujący z systemami wielocelowymi wie, że nie jest to elegancka matematyczna zabawa, lecz codzienny dramat kombinatoryczny. Im gęstszy ruch i większa niepewność pomiarowa, tym szybciej system zbliża się do nieuchronnej ściany wydajności. A ściana, jak uczy doświadczenie, jest najbardziej realistycznym uczestnikiem większości ambitnych projektów technologicznych.

Właśnie tutaj adiabatyczne obliczenia kwantowe oraz mapowanie problemów na model Isinga ujawniają swoją rzeczywistą wagę. Istota tej koncepcji jest zarazem subtelna i brutalnie praktyczna, gdyż zmienia samo paradygmatyczne podejście do poszukiwania rozwiązań. Zamiast zmuszać klasyczny procesor do żmudnego przeszukiwania gigantycznej przestrzeni, kodujemy funkcję

kosztu jako hamiltonian konkretnego problemu. System ewoluuje wówczas naturalnie ku stanowi najniższej energii, który odpowiada rozwiązaniu optymalnemu lub przynajmniej bliskiemu ideałowi. Jest to głęboka zmiana, w której przestajemy traktować maszynę jako cyfrowego biurokratę, a zaczynamy widzieć w niej układ fizyczny. Pozwalamy naturze rozstrzygać, które konfiguracje danych są stabilne, a które stanowią jedynie nieistotny szum.

Nie każemy już algorytmowi kolejno rozważać każdej kandydatury, lecz przenosimy ciężar z logiki enumeracyjnej do logiki czysto energetycznej. Z perspektywy filozofii techniki jest to posunięcie niemal klasyczne, wpisujące się w odwieczne dążenie do delegowania trudnych zadań samej naturze. Człowiek od dawna próbuje outsourcować te rachunki, których jego własny aparat poznawczy nie jest w stanie wykonać dostatecznie szybko. Silnik parowy zastąpił pracę mięśni, radar przejął fragment ludzkiej percepcji, a systemy kwantowe delegują optymalizację na poziom fizycznej ewolucji układu. W przeciwnym razie koncepcja „human in the loop" staje się jedynie liturgią odpowiedzialności bez realnej sprawczości. Odpowiada człowiek. Zawsze człowiek.

## Etyka i wyjaśnialność jako fundamenty bezpieczeństwa AI

Zachowajmy chłód analityczny, bowiem adiabatyczne obliczenia kwantowe nie są uniwersalnym lekarstwem na każdą bolączkę asocjacji danych. Ich realna przydatność w systemach obronnych zależy od brutalnej fizyki: skali układu, jakości implementacji, odporności na szum oraz całej ekonomii urządzenia, która musi zmieścić się w ramach luki energetycznej. Nie zamierzam tu pisać hymnu na cześć kwantowego zbawienia, gdyż technologia ta, choć fascynująca, pozostaje zakładnikiem swoich materialnych ograniczeń. Chodzi raczej o dostrzeżenie fundamentalnej zmiany mentalnej, w której algorytm przestaje być tylko abstrakcyjnym przepisem, a staje się procesem głęboko zakorzenionym w materii. Komputacja nie jest już wyłącznie procedurą operującą na symbolach, lecz precyzyjnie zaprojektowanym sposobem wykorzystania właściwości fizycznych do rozwiązywania problemów informacyjnych.

Tutaj powraca motyw ekonomiczny, bo problemy asocjacji danych czy selekcji hipotez są kosztowne nie tylko w sensie czysto obliczeniowym. Generują one potężne koszty instytucjonalne, gdzie każde opóźnienie, błąd identyfikacji czy fałszywy alarm pociąga za sobą lawinę wydatków wtórnych. Mówimy o kosztach decyzji spóźnionych, błędnych, o przeciążeniu operatora i wreszcie o tragicznych w skutkach szkodach ubocznych, których nie da się wycenić w prostym arkuszu kalkulacyjnym. W tym ujęciu techniki kwantowe i zaawansowane heurystyki nie są jedynie sztuką przyspieszania pracy procesora, lecz próbą obniżenia strategicznego kosztu niepewności. Wdrażana

tu architektura selekcji, czyli modelowanie AI jako systemu aktywnie filtrującego sygnały i priorytety, pozwala na realne zarządzanie tymi kosztami w warunkach bojowych.

Pole walki czy infrastruktura monitorująca to nie są sterylne laboratoria dla czystych pojęć, lecz skomplikowane układy, w których każda poprawa jakości rozstrzygnięcia ma wymierne konsekwencje. Właśnie dlatego wyjaśnialność AI, znana jako xAI, musi stać w samym centrum systemów bezpieczeństwa, a nie na ich dekoracyjnym marginesie obok haseł o innowacyjności. Im bardziej złożone stają się mechanizmy predykcji, tym mniej wystarcza nam sama trafność wyjściowa, bo bez zrozumienia procesu decyzyjnego jesteśmy skazani na ślepe zaufanie. Kluczowa staje się tutaj aksjologia progu alarmowego, czyli uznanie, że wybór czułości sensora jest decyzją wartościującą, a nie tylko techniczną. Potrzebujemy zdolności przedstawienia człowiekowi, na jakiej konkretnie podstawie system osiągnął dany wniosek, co stanowi fundament bezpiecznego zarządzania ryzykiem.

NIST już w 2021 roku sformułował cztery zasady wyjaśnialnej AI, podkreślając, że system musi dostarczać wyjaśnień znaczących dla użytkownika i poprawnych w granicach swoich założeń. To niezwykle istotne, ponieważ w potocznym obiegu xAI bywa często redukowana do estetycznego komentarza, swoistego dashboardu, który ma jedynie uspokoić sumienie operatora. Tymczasem chodzi o coś znacznie twardszego: o możliwość krytycznej oceny, czy wynik opiera się na sensownych przesłankach, czy może na przypadkowych korelatach i artefaktach. Komputer, jak każdy współczesny biurokrata, bywa szczególnie groźny wtedy, gdy myli pewność z kompetencją, serwując nam błędy ubrane w szaty matematycznej nieomylności. Musimy przy tym uważać na automation bias, czyli tendencję operatorów do bezkrytycznego ufania algorytmom, co prowadzi do niebezpiecznej, fikcyjnej delegacji odpowiedzialności.

W wojskowym użyciu AI ten problem staje się palący, co doskonale widać na przykładzie systemów wspierających automatyczne rozpoznawanie celów oraz szacowanie szkód pobocznych. W takich obszarach nie wystarczy wysoka skuteczność laboratoryjna, gdyż liczy się przede wszystkim to, czy operator pod presją czasu zrozumie podstawy otrzymanej rekomendacji. Musimy wiedzieć, czy stosowana analiza cech istotnych rzeczywiście zwiększa zdolność krytycznego sądu, czy jedynie podaje nam iluzję uzasadnienia, maskującą brak realnej kontroli. To rozróżnienie jest kluczowe, bo bez zachowania meaningful human control, czyli standardu nadzoru dającego realną możliwość interwencji, system staje się samowolny. W przeciwnym razie human in the loop staje się liturgią odpowiedzialności bez odpowiedzialności, gdzie człowiek jedynie przyklepuje wyroki maszyny.

Współczesne metody xAI często dają jedynie wgląd lokalny, nie potrafiąc zrekonstruować rzeczywistych związków przyczynowych w skali globalnej, co wymusza na nas dużą dozę ostrożności. Dzisiejszy paradygmat naukowy mówi raczej o użytecznych narzędziach interpretacji

niż o ostatecznym rozwiązaniu problemu „czarnej skrzynki”, co jest podejściem uczciwym i bezpiecznym. Komisja Europejska, projektując nowe ramy regulacyjne, również nie sugeruje, że sama przejrzystość znosi ryzyko, lecz widzi w niej sposób na wzmocnienie nadzoru i rozliczalności. Tutaj dochodzimy do etyki w sensie ścisłym, gdzie trzy wielkie tradycje normatywne – konsekwencjalizm, deontologia i etyka cnót – przestają być tylko akademicką dekoracją. Każda z nich wnosi unikalną perspektywę do namysłu nad systemami sensorowymi, które coraz częściej decydują o ludzkim losie.

Konsekwencjalizm przypomina nam, że nie wolno oceniać systemu bez odniesienia do jego realnych skutków, w tym do asymetrycznego rozkładu ryzyka i błędów identyfikacji. Z kolei deontologia stawia twarde granice, przypominając, że istnieją normy i procedury, których nie można złamać nawet w imię najwyższej sumy korzyści operacyjnych. Najciekawsza wydaje się jednak etyka cnót, która podkreśla, że nawet najdoskonalszy algorytm nie zdejmie z człowieka obowiązku roztropności, odwagi oraz praktycznej mądrości. Można zautomatyzować tysiące operacji i procesów, ale nie da się zautomatyzować charakteru ani poczucia odpowiedzialności za podjęte działania. Dobry system wie, że nim nie jest.

Właśnie dlatego teza, że odpowiedzialność należy wyłącznie do ludzi, pozostaje nie tylko poprawna filozoficznie, lecz także absolutnie konieczna z punktu widzenia politycznego. Maszyna nie może być podmiotem moralnym; może być jedynie źle zaprojektowana, niewłaściwie wytrenowana lub błędnie włączona w procesy decyzyjne przez swoich twórców. Odpowiadają zawsze ludzie i instytucje, co znajduje swoje odzwierciedlenie w unijnym AI Act, który wszedł w życie 1 sierpnia 2024 roku. Rozporządzenie to wprowadza rygorystyczne wymogi dla dostawców systemów wysokiego ryzyka, traktując technologię jako narzędzie, a nie jako autonomiczny podmiot prawa. Taka architektura prawna wymusza na nas odejście od outsourcingu sumienia na rzecz świadomego projektowania systemów, które wspierają, a nie zastępują ludzki osąd.

## Architektura odpowiedzialności: między kontrolą a alibi

Współczesne państwo, wbrew technologicznym utopiom, rzadko daje się uwieść opowieściom o moralnej samodzielności maszyn. Zamiast tego, z niemal urzędniczą skrupulatnością, próbuje rozdzielić role, obowiązki dokumentacyjne oraz nadzorcze na te sprawowane przed i po fakcie. Niemniej jednak pod tą warstwą proceduralnego spokoju kryje się zjawisko znacznie bardziej niepokojące dla naszej podmiotowości. Im doskonalsza staje się sztuczna inteligencja, tym silniejsza okazuje się pokusa fikcyjnej delegacji odpowiedzialności na algorytm. Człowiek, choć

formalnie nadal podpisuje każdą kluczową decyzję, w praktyce coraz rzadziej ją merytorycznie konstytuuje, stając się jedynie zakładnikiem procesora.

To właśnie tutaj objawia się prawdziwy dramat zjawiska automation bias, czyli tendencji operatorów do bezkrytycznego ufania wynikom generowanym przez algorytmy, nawet przy ich ewidentnej błędności. Nie chodzi tu wyłącznie o to, że operator ufa systemowi zbyt mocno, lecz o to, że cała architektura poznawcza sytuacji zostaje ułożona przeciwko jego nieufności. Wynik modelu przychodzi błyskawicznie, jest wsparty sugestywną wizualizacją i opatrzony wysokim wskaźnikiem pewności. Często towarzyszy mu dodatkowo gotowa interpretacja, która zdejmuje z barków użytkownika ciężar samodzielnego myślenia. W takim układzie sceptycyzm staje się po prostu niefunkcjonalny, a każda próba weryfikacji jest postrzegana jako zbędny koszt.

W efekcie człowiek otrzymuje dramatycznie mało czasu, przytłaczającą odpowiedzialność i coraz gorsze warunki do realnej odmowy. Wtedy human in the loop, czyli koncepcja zakładająca obecność człowieka w pętli decyzyjnej, staje się jedynie formą liturgii odpowiedzialności bez odpowiedzialności. Jest to rytualny udział w procesie, który został już uprzednio rozstrzygnięty przez maszynową selekcję sygnałów i priorytetów. Dlatego tak istotny jest postulat meaningful human control, rozumiany jako standard nadzoru gwarantujący realną, a nie tylko ceremonialną możliwość interwencji. Kontrola ta nie może być przecież wyłącznie obecnością nazwiska na schemacie architektury systemu.

Prawdziwe sprawstwo musi obejmować głębokie rozumienie procesu, czas na reakcję oraz kompetencję do zakwestionowania modelu w dowolnym momencie. Wymaga to istnienia procedur eskalacji i, co najważniejsze, kultury organizacyjnej, która nie karze podwładnych za sprzeciw wobec rekomendacji systemu. W tym punkcie rozważania o algorytmicznym zarządzaniu i godności okazują się zaskakująco ważne dla świata nowoczesnej obronności. Artykuł przypomina nam bowiem, że technologia nie jest ślepym żywiołem, lecz skutkiem konkretnych decyzji inwestycyjnych oraz politycznych. Prawo może i powinno działać jako narzędzie kształtowania realnych granic władzy algorytmicznej nad życiem.

Nie istnieje żadna metafizyczna konieczność, która zmuszałaby państwa, armie czy korporacje do projektowania systemów odbierających ludziom rzeczywiste sprawstwo. To zawsze jest wybór organizacyjny, a marketing, jak wiadomo, potrafi sprzedać nawet mgłę jako infrastrukturę krytyczną, byle tylko nadać jej efektowny dashboard. Można budować systemy wspierające decyzję, ale można też tworzyć mechanizmy dyscyplinujące użytkownika i ukrywające przesłanki działania. Te drugie maksymalizują szybkość kosztem refleksji, udając przy tym, że odpowiedzialność nadal spoczywa na barkach człowieka. Komputer, jak każdy współczesny biurokrata, bywa szczególnie groźny wtedy, gdy myli pewność z kompetencją.

Najbardziej etyczny system AI to nie taki, który rzekomo posiada moralną duszę i sam odróżnia dobro od zła. Taka wizja jest zarezerwowana dla naiwnych entuzjastów lub działów PR, które chcą ukryć techniczne niedoskonałości pod płaszczem humanizmu. Prawdziwie etyczna technologia cechuje się instytucjonalną pokorą, nie ukrywając przed człowiekiem struktury podjętej decyzji i nie rozpuszczając odpowiedzialności w automatyzacji. Nie wymusza ona bezrefleksyjnej zgody i nie odcina możliwości zakwestionowania wygenerowanej rekomendacji. Dobry system po prostu wie, że nie jest człowiekiem i nie usiłuje go udawać w procesie moralnym.

W tym sensie cała współczesna technologia, od klasyfikacji sygnałów po obliczenia kwantowe, daje się sprowadzić do jednej osi cywilizacyjnej. AI nie jest jedynie narzędziem automatyzacji, lecz technologią zarządzania niepewnością, która porządkuje chaos tam, gdzie ludzka percepcja przestaje wystarczać. Klasyfikuje ona słabe sygnały, przewiduje trajektorie i alokuje naszą ograniczoną uwagę w sposób niezwykle efektywny. Wykorzystuje przy tym redukcję złożoności, czyli proces brutalnego upraszczania rzeczywistości w celu umożliwienia szybkiego działania operacyjnego. Kto lepiej zarządza tą niepewnością, ten wcześniej reaguje i wcześniej ustanawia własną wersję rzeczywistości.

Ostatecznie spór o sztuczną inteligencję nie jest sporem o parametry techniczne modeli, lecz o architekturę odpowiedzialności w świecie rosnącej mocy obliczeniowej. Maszyna może policzyć nieskończenie więcej niż my, ale nie może sama odpowiedzieć na pytanie, czy to, co policzyła, powinno zostać wykonane. I całe szczęście, bo jeśli kiedyś uwierzymy, że potrafi ona przejąć nasz ciężar etyczny, to nie wkroczymy do epoki rozumu maszyn. Wkroczymy w mroczne czasy ludzkiego alibi, gdzie każda decyzja będzie jedynie pochodną algorytmu, za który nikt nie chce już brać osobistej odpowiedzialności.

## Zarządzanie zasobami jako konstytucja systemu i polityka

W samym sercu zagadnienia nie odnajdziemy kwantowej elegancji ani kunsztu klasyfikacji, lecz prozaiczne, a zarazem fundamentalne zarządzanie zasobami. To właśnie tutaj teoria przestaje być widowiskiem dla intelektualistów, a staje się twardą administracją realności, gdzie każda mikrosekunda ma swoją wymierną cenę. Współczesny wielofunkcyjny radar nie jest już prostym urządzeniem wykonującym jedną czynność, lecz polem nieustannej, wewnętrznej konkurencji o priorytety. Śledzenie domaga się częstego odświeżania danych, przeszukiwanie żąda szerokiego pokrycia przestrzeni, a identyfikacja potrzebuje czasu i wysokiej jakości sygnału. Każda z tych

funkcji zgłasza własne, agresywne roszczenie do wspólnego budżetu czasu, energii oraz mocy obliczeniowej systemu.

Dlatego zarządzanie zasobami (Resource Management), czyli dynamiczna administracja czasem i mocą obliczeniową sensora w warunkach wewnętrznej konkurencji zadań, nie jest jedynie technicznym detalem implementacyjnym. Stanowi ono faktyczną konstytucję systemu, która rozstrzyga w ułamku sekundy, co radar uzna za warte swojej uwagi, swojej wiązki i swojego operacyjnego ryzyka. W tym sensie zarządzanie zasobami jest polityką pod inną nazwą, a jakość obsługi (Quality of Service) stanowi jej ekonomię normatywną. Próbuje ona bowiem przeliczyć jakości całkowicie niewspółmierne na wspólną, matematyczną skalę użyteczności, co zawsze wiąże się z arbitralnym wyborem wartości. Źródłowy materiał trafnie ujmuje ten proces jako przejście od technicznych metryk ku jednolitej logice użyteczności (utility).

Współczesne debaty wokół wdrażania sztucznej inteligencji podkreślają, że projekt pozbawiony spójnej architektury celów staje się kosztownym chaosem zamiast systemem wspierającym decyzje. Wątek z portalu dobrepanstwo.org jest tu szczególnie cenny, gdyż fundacja nie traktuje AI jak meteoru spadającego na instytucje. Zamiast tego postuluje osadzenie technologii w szkielecie organizacyjnym, wskaźnikach i kryteriach odpowiedzialności, które nadają jej sens. Artykuł o architekturze innowacji akcentuje, że każdy projekt musi posiadać uzasadnienie biznesowe oraz jasne miejsce w szerszej strategii rozwoju. A marketing, jak wiadomo, potrafi sprzedać nawet mgłę jako infrastrukturę krytyczną, byle tylko nadać jej efektowny pulpit nawigacyjny (dashboard).

Przeniesione na grunt sensorów i nowoczesnej obronności oznacza to rzecz bardzo prostą, choć dla wielu inżynierów wyjątkowo niewygodną. Nie ma sensu twierdzić, że system wielosensorowy został zoptymalizowany, jeśli nie zdefiniowano wcześniej, dla jakiej teorii sukcesu dokonano tego strojenia. Algorytm, który maksymalizuje źle ustawioną użyteczność, może działać technicznie bezbłędnie i zarazem systematycznie szkodzić całości powierzonej mu misji. Innymi słowy, QoS nie jest apolityczną matematyką dobrobytu systemu, lecz kodowaniem preferencji instytucji wbudowanym w mechanizm alokacji. Jeśli wartości wpisane do funkcji celu są płytkie, wskaźniki staną się tylko eleganckim narzędziem reprodukcji tego spłycenia.

Warunki optymalności, które w literaturze przedmiotu pojawiają się pod postacią formalizmu KKT, można oczywiście odczytywać jako neutralny aparat matematyczny. Jednak dla eksperta znacznie bardziej interesujące jest to, co kryje się pod tą fasadą technicznej bezstronności. Osiągnięcie punktu, w którym żadne dalsze przesunięcie zasobów nie zwiększa już łącznej użyteczności, brzmi jak ostateczny triumf czystej racjonalności. Dopóki jednak nie zapytamy, jak skonstruowano funkcję użyteczności, nie wiemy, czy mówimy o mądrości, czy o sprawnej metodzie konserwowania

cudzych założeń. To odwieczny problem ekonomii publicznej, gdzie system potrafi doskonale optymalizować to, co zostało fundamentalnie źle zdefiniowane.

W radarze oznacza to możliwość perfekcyjnego gospodarowania czasem anteny w ramach błędnie ustawionej hierarchii celów operacyjnych. W państwie natomiast oznacza to możliwość wydajnego administrowania procesami, które w ogóle nie powinny być w ten sposób prowadzone. W obu przypadkach dostrzegamy ten sam mechanizm: formalna poprawność nigdy nie zastąpi normatywnej mądrości decydenta. I właśnie dlatego etyka oraz zarządzanie zasobami nie są dwoma odrębnymi rozdziałami podręcznika, lecz dwiema stronami tej samej decyzji. Rozstrzygają one, co w warunkach nieuchronnego niedoboru zasługuje na pierwszeństwo, a co musi zostać pominięte.

Współczesna recepcja naukowa idzie dokładnie w tym kierunku, odchodząc od modeli czysto laboratoryjnych ku twardej rzeczywistości wdrożeniowej. Coraz mniej przekonuje model, w którym ocena systemu AI kończy się na wskaźnikach skuteczności uzyskanych w sterylnych warunkach. NIST zwraca uwagę, że wdrożona sztuczna inteligencja (deployed AI) wymaga ciągłego monitoringu, ponieważ rzeczywisty świat nie zachowuje się jak zamknięty wzorzec (benchmark). W systemach sensorowych to zastrzeżenie jest wręcz oczywiste, gdyż model klasyfikacji sygnału może zawodzić po najmniejszej zmianie profilu zakłóceń. Rzeczywistość operacyjna to nieustanna zmiana emitera, reorganizacja ruchu obiektów czy celowa manipulacja ze strony przeciwnika.

Mechanizm alokacji zasobów, który sprawdzał się przy pewnym poziomie zagęszczenia śladów, może okazać się całkowicie zawodny przy innej strukturze misji. To oznacza, że QoS oraz zarządzanie zasobami nie są zagadnieniami jednorazowego strojenia, lecz procesami trwałego zarządzania niepewnością instytucjonalną. Stąd bierze się rosnące znaczenie certyfikacji oraz procedur audytowych, które chronią system przed manipulacją i błędami operatorów. To nie jest biurokratyczny naddatek do prawdziwej inżynierii, lecz integralna część samej inżynierii odpowiedzialności. W systemach wysokiego ryzyka nie wystarcza, że model działa; trzeba jeszcze wykazać, jak i w jakim zakresie to robi.

Właśnie w tym kierunku zmierza europejskie podejście regulacyjne, które przestało traktować AI jako neutralną nowinkę technologiczną. Komisja Europejska przypomina, że AI Act wszedł w życie 1 sierpnia 2024 roku, a jego stosowanie następuje w ściśle określonych etapach. Zakazane praktyki oraz obowiązki dotyczące kompetencji w zakresie AI zaczęły obowiązywać już od 2 lutego 2025 roku. Kolejne kroki obejmują modele ogólnego przeznaczenia oraz pełną stosowalność aktu, która ma nastąpić w sierpniu 2026 roku. Ten harmonogram pokazuje, że Zachód zaczął traktować sztuczną inteligencję jako istotny obiekt zarządzania ustrojowego i prawnego.

Dla profesjonalnego czytelnika szczególnie ważne jest jednak to, czego żadna, nawet najlepsza regulacja nie może zrobić za nas. Prawo może wymusić dokumentację, obowiązki dostawcy, procedury nadzoru oraz wymagania przejrzystości, ale nie zastąpi ono znaczącej kontroli ludzkiej (Meaningful Human Control). Jest to standard nadzoru, w którym człowiek posiada realną, a nie tylko ceremonialną możliwość interwencji w działania systemu. Zapobiega to zjawisku stronniczości automatyzacji (Automation Bias), czyli tendencji operatorów do bezkrytycznego ufania wynikom algorytmów, nawet przy ich ewidentnej błędności. W przeciwnym razie człowiek w pętli (human in the loop) staje się liturgią odpowiedzialności bez odpowiedzialności, gdzie człowiek jedynie podpisuje decyzje, których nie rozumie.

## Pułapka pozornej kontroli: między alibi a sprawstwem

Nie zastąpi to jednak rzetelnego namysłu nad tym, jaka architektura kontroli człowieka jest realna, a jaka jedynie ceremonialna. W gąszczu systemów wspierających decyzje operator często przypomina ducha w maszynie – jest formalnie obecny, lecz poznawczo całkowicie nieobecny. Może on mechanicznie klikać zatwierdzenie, nie mając w rzeczywistości żadnej możliwości odrzucenia maszynowej rekomendacji. Może wpatrywać się w mapę cieplną, czyli graficzną reprezentację istotności danych, ale nie posiadać czasu, by ocenić, czy wskazuje ona trafną cechę, czy tylko cyfrowy artefakt. To właśnie największa pokusa nowoczesnej automatyzacji: eleganckie rozdzielenie odpowiedzialności formalnej od sprawstwa materialnego.

Źródłowa teza o konieczności znaczącej kontroli człowieka (Meaningful Human Control), czyli standardu nadzoru, w którym człowiek posiada realną możliwość interwencji, pozostaje więc warunkiem zachowania sensu odpowiedzialności prawnej i moralnej. Bez tego fundamentu cały system staje się jedynie wyrafinowaną machiną do produkowania ludzkiego alibi dla maszynowej selekcji. W świecie współczesnych organizacji łatwo wyobrazić sobie idealnie nowoczesny system, który posiada pulpity nawigacyjne, modele generatywne i panel z logo w odcieniu szlachetnego granatu, a mimo to działa jak automat do produkcji pozorów wiedzy. A marketing, jak wiadomo, potrafi sprzedać nawet mgłę jako infrastrukturę krytyczną, byle tylko nadać jej odpowiedni interfejs.

Taki system wysyła do operatora uspokajający komunikat: proszę się nie martwić, wszystko jest pod kontrolą, a w razie katastrofy człowiek pozostaje w pętli. I rzeczywiście, człowiek w tej pętli pozostaje, choć najczęściej w tej jej części, która musi podpisać raport po wystąpieniu incydentu. W przeciwnym razie człowiek w pętli (human in the loop) staje się liturgią odpowiedzialności bez odpowiedzialności, rytuałem odprawianym ku uspokojeniu sumień regulatorów. Absurd polega na

tym, że im bardziej złożone są systemy, tym silniej kuszą instytucje, by mylić dekorację kontroli z jej faktycznym sprawowaniem.

Dlatego trzeba uporczywie powtarzać, że xAI, czyli wyjaśnialna sztuczna inteligencja, nie jest relikwią epistemologiczną do oglądania przez szybę, lecz instrumentem przywracania prawa do sceptycyzmu. Jeżeli narzędzie to nie zwiększa realnej zdolności zakwestionowania modelu, staje się jedynie lepiej podświetlonym interfejsem posłuszeństwa. Współczesny paradygmat naukowy zaczyna to rozumieć coraz lepiej, akcentując, że interpretowalność jest zawsze zależna od kontekstu użycia. Jakość wyjaśnienia nie może być utożsamiana z samą jego obecnością, bo martwe dane rzadko budują autentyczne zrozumienie.

NIST słusznie podkreśla, że wyjaśnienie musi być znaczące dla konkretnego użytkownika, co w praktyce oznacza radykalne zróżnicowanie komunikatów. Inne wyjaśnienie ma wartość dla inżyniera debugującego model, inne dla operatora działającego pod presją czasu, a jeszcze inne dla audytora czy organu nadzoru. To rozróżnienie jest krytyczne dla systemów bojowych i sensorowych, gdzie abstrakcyjna wyjaśnialność jest całkowicie bezużyteczna. Tam system musi być zrozumiały w czasie, tempie i języku decyzji, które rzeczywiście zapadają na polu walki.

Inaczej nawet doskonały model interpretacyjny pozostaje bezużyteczny dla praktyki, podobnie jak mądra mapa dla człowieka spadającego z urwiska. Taki pechowiec nie potrzebuje wykładu z geologii klifu, lecz wiedzy o tym, gdzie postawić stopę w następnej sekundzie. Tendencja do bezkrytycznego ufania algorytmom (Automation Bias) karmi się właśnie takimi lukami między teorią a operacyjną rzeczywistością. Z tej perspektywy konstrukcja artykułu zaczyna się domykać, wskazując, że prawdziwa kontrola wymaga czegoś więcej niż tylko obecności przycisku „stop”.

## AI jako architektura selekcji: między techniką a ustrojem

Klasyfikacja sygnałów w nowoczesnych systemach obronnych przestała być prostym ćwiczeniem z rozpoznawania wzorców, stając się w istocie głęboką organizacją ludzkiej percepcji. Współczesny wywiad elektroniczny, znany jako ELINT, coraz mniej przypomina katalog motyli, a coraz bardziej hermeneutykę maszyn, czyli sztukę interpretacji zachowań urządzeń w gęstym szumie informacyjnym. Nie chodzi tu o zwykłe gromadzenie danych, lecz o nadawanie sensu emisjom, które dla nieuzbrojonego oka pozostają jedynie chaosem. W tym procesie technologia TBD, czyli Track before Detect, nie jest wyłącznie sprytnym obejściem progów detekcji, lecz kulturą akumulowania słabej wiedzy wbrew pokusie przedwczesnej pewności. System uczy się cierpliwości,

zbierając drobne okruchy informacji, zanim odważy się ogłosić istnienie celu, co fundamentalnie zmienia naszą definicję dowodu. Ostatecznie rozpoznanie sygnałów staje się laboratorium trudnych kompromisów między wykrywalnością a ryzykiem fałszywego alarmu, gdzie stawką jest godność i bezpieczeństwo całych społeczeństw.

Zarządzanie zasobami, określane mianem Resource Management, nie jest w tym kontekście jedynie nudnym detalem wydajnościowym, lecz ekonomią niedoboru przetłumaczoną na konkretne decyzje maszynowe. Kiedy sensor musi wybierać między śledzeniem pocisku a monitorowaniem przestrzeni powietrznej, dokonuje on wyboru o charakterze niemal politycznym, ustalając hierarchię zagrożeń. Podobnie wskaźnik QoS, czyli Quality of Service, nie jest niewinnym parametrem technicznym, lecz formalizacją instytucjonalnych preferencji, które określają, co w danej chwili jest dla państwa priorytetem. Aksjologia progu alarmowego, czyli uznanie, że wybór czułości sensora jest decyzją wartościującą, uświadamia nam, że każdy system detekcji opiera się na kompromisie między bezpieczeństwem a ryzykiem. Zasada wykluczenia i obliczenia kwantowe przestają być wówczas jedynie ornamentem nowoczesności, stając się próbą ujarzmienia problemów nierozróżnialności przy pomocy zaawansowanego aparatu fizycznego. To tutaj xAI, czyli wyjaśnialna sztuczna inteligencja, przestaje być kosmetyką zaufania, a staje się warunkiem zachowania resztek suwerenności człowieka wobec systemu, który widzi więcej i szybciej.

Etyka w projektowaniu systemów obronnych nie powinna być traktowana jako końcowy, obowiązkowy rozdział raportu, lecz jako fundamentalne pytanie o sens odpowiedzialności w dobie automatyzacji. Musimy zastanowić się, czy w ogóle potrafimy jeszcze projektować technikę tak, by nie unieważniała ona podmiotowości operatora i nie zamieniała go w bezwolny element pętli decyzyjnej. Takie podejście dobrze współbrzmi z linią intelektualną obecną na portalu dobrepanstwo.org, gdzie technologia jest zawsze osadzona w strukturze państwa, prawa i godności człowieka. W tekstach Fundacji powraca słuszne przekonanie, że narzędzia obliczeniowe nie są neutralne, ponieważ modelują one relacje zależności oraz hierarchie wiedzy w nowoczesnym społeczeństwie. Kto projektuje system selekcji sygnałów i alarmów, ten w istocie projektuje pewną antropologię polityczną, definiując, komu ufać i co uznać za dopuszczalne odchylenie od normy. To są pytania stare jak państwo i wojna, które AI jedynie wyostrza, przyspiesza i bezlitośnie zamyka w linijkach kodu, czyniąc je trudniejszymi do dostrzeżenia.

Najuczciwsza konkluzja dotycząca współczesnych zmian nie brzmi wcale, że nadchodzi epoka autonomicznych maszyn, gdyż taka formuła jest zbyt łatwa i intelektualnie tania. Nadchodzi raczej czas systemów, które coraz skuteczniej zarządzają niepewnością, lecz właśnie przez to obnażają niedojrzałość naszych instytucji publicznych i procedur nadzoru. Jeśli państwo próbuje przykryć brak odpowiedzialnej architektury samym zachwytem nad wydajnością algorytmów, ryzykuje

utratę kontroli nad kluczowymi procesami decyzyjnymi. A marketing, jak wiadomo, potrafi sprzedać nawet mgłę jako infrastrukturę krytyczną, byle tylko nadać jej dashboard i kilka efektownych wykresów. Przyszłość należy zatem nie do tych, którzy zbudują najbardziej efektowny model, lecz do tych, którzy potrafią związać go z prawem, audytem i rzetelnym treningiem operatorów. Innymi słowy, przetrwają te systemy, które potrafią uczynić sztuczną inteligencję rozliczalną przed społeczeństwem, zamiast tworzyć technokratyczne baśnie o nieomylności maszyn.

Sztuczna inteligencja w środowisku sensorowym i analitycznym nie jest przede wszystkim techniką imitowania ludzkiego rozumu, lecz wyrafinowaną metodą organizowania niepewności w świecie nadmiaru danych. Klasyfikuje ona sygnały w momentach, gdy ludzkie ucho słyszy już tylko jednostajny szum, a oko nie dostrzega żadnych istotnych zmian w obserwowanym krajobrazie. Wiąże ślady i hipotezy wtedy, gdy przestrzeń obserwacji zaczyna rozpadać się na konkurencyjne, często sprzeczne ze sobą scenariusze wydarzeń, wymagające natychmiastowej reakcji. Przewiduje trajektorie w sytuacjach, gdy modele liniowe przestają wystarczać, a dynamika pola walki wymyka się prostym schematom myślowym i intuicji. Wskazuje ona, co uznać za podejrzane, co za dopuszczalne, a co za akceptowalny koszt błędu, co w praktyce oznacza brutalną redukcję złożoności świata. Właśnie dlatego jej rdzeniem nie jest banalna automatyzacja, lecz selekcja, która natychmiast staje się problemem ustrojowym, a nie wyłącznie technicznym wyzwaniem dla inżynierów.

Warto przyjrzeć się bliżej, dlaczego klasyfikacja sygnałów akustycznych czy radarowych jest tak kluczowym zagadnieniem dla współczesnej architektury bezpieczeństwa i stabilności państwa. W klasyfikacji radarowej pojedynczy impuls przestaje wystarczać, ponieważ nowoczesny emiter zdradza się nie przez jeden parametr, lecz przez własną, unikalną gramatykę działania. Wymusza to przejście od prostego katalogu cech do pogłębionej analizy zachowań, co wymaga od systemów ogromnej mocy obliczeniowej i precyzji w interpretacji. W rozpoznaniu otwartego zbioru oznacza to obowiązek epistemicznej pokory, czyli zdolność systemu do przyznania, że napotkał zjawisko, którego po prostu nie zna i nie potrafi sklasyfikować. Dobrze zaprojektowany klasyfikator nie tylko rozpoznaje wzorce, ale także wie, kiedy nie powinien udawać wiedzy, której mu w danej chwili brakuje. Komputer, jak każdy współczesny biurokrata, bywa szczególnie groźny wtedy, gdy myli pewność z kompetencją, prowadząc nas na manowce błędnych założeń.

NIST ujmuje to jako zasadę knowledge limits, czyli konieczność respektowania granic własnego zakresu działania, co w obszarach wysokiego ryzyka jest cnotą fundamentalną i rzadką. Model, który nie umie powiedzieć „nie wiem”, zbyt łatwo staje się maszyną do produkowania kosztownej pewności pozornej, co może prowadzić do tragicznych w skutkach decyzji operacyjnych. Wspomniana technika Track before Detect nie jest zatem zwykłą poprawką do starej architektury

wykrywania, lecz fundamentalną zmianą filozofii poznania i podejścia do danych. W miejsce logiki jednego silnego sygnału wprowadza ona logikę akumulacji słabych przesłanek, które dopiero razem tworzą spójny i wiarygodny obraz rzeczywistości. Zamiast progu jako binarnej bramy prawdy, otrzymujemy proces stopniowego wyłaniania się obiektu z tła, co ma kolosalne znaczenie w świecie małych dronów i pasywnej radiolokacji. Ostatecznie każda taka innowacja techniczna niesie ze sobą głębokie konsekwencje dla sposobu, w jaki rozumiemy nadzór, suwerenność oraz Meaningful Human Control. Ten standard nadzoru, w którym człowiek posiada realną możliwość interwencji w działania systemu, staje się jedyną barierą przed całkowitą alienacją technologii.

## Architektura wyboru: Dlaczego AI to polityka, nie technologia

W systemach krytycznych prawda operacyjna coraz rzadziej objawia się jako nagły błysk intuicji, stając się raczej żmudnie kondensowanym rozkładem prawdopodobieństwa. Kto pragnie dostrzec zagrożenie wcześniej niż inni, musi zapłacić za to wysoką cenę w walucie mocy obliczeniowej, pamięci oraz cierpliwości wobec nieuniknionej niejednoznaczności. Wszak sztuczna inteligencja wcale nie usuwa niepewności z pola walki, lecz jedynie uczy instytucje, jak kupować z nią rozejm na nieco korzystniejszych warunkach. Jest to opis znacznie bardziej realistyczny i naukowo uczciwszy niż modne, marketingowe bajki o rzekomej wszechwiedzy wielkich modeli analitycznych. Redukcja złożoności, czyli proces brutalnego upraszczania rzeczywistości przez algorytmy w celu umożliwienia szybkiego działania, zawsze odbywa się kosztem pewnych pominiętych detali. Niemniej to właśnie ta selektywność pozwala nam na reakcję w czasie, który wciąż daje szansę na przetrwanie.

Jeszcze ostrzej ta zależność ujawnia się w obszarze rozpoznania nuklearnego, gdzie margines błędu praktycznie nie istnieje, a stawka jest ostateczna. Tam fuzja danych nie jest technologicznym luksusem dla entuzjastów, lecz absolutną koniecznością operacyjną, bez której system pozostaje ślepy. Sam detektor promieniowania gamma nie wystarczy do podjęcia decyzji, podobnie jak analiza samej trajektorii ruchu obiektu rzadko bywa w pełni konkluzywna. Dopiero ich ścisłe sprzężenie, a następnie probabilistyczne modelowanie w czasie, pozwala uzyskać coś, co możemy nazwać praktycznym rozpoznaniem sytuacji. Przy czym wraz z tym wzrostem technicznej skuteczności rośnie również ciężar normatywny, który spoczywa na barkach projektantów i decydentów. Każdy próg alarmu oraz każda decyzja o czułości systemu stają się w tym kontekście deklaracją polityczną.

Aksjologia progu alarmowego, czyli uznanie, że wybór czułości sensora jest decyzją wartościującą, uświadamia nam, iż technologia nigdy nie jest etycznie neutralna. Każdy kompromis między wysoką wykrywalnością a akceptowalną liczbą fałszywych alarmów jest w istocie wyborem moralnym, który rozstrzyga o ludzkim życiu. Komisja Europejska, opisując AI Act, stale podkreśla, że celem ram regulacyjnych jest pogodzenie innowacji z ochroną zdrowia i praw podstawowych. Nie jest to bynajmniej jałowa biurokratyczna ornamentyka, lecz próba okiełznania sił, które mogą wymknąć się spod kontroli. To formalne uznanie faktu, że systemy wysokiego ryzyka zawsze organizują czyjąś ekspozycję na błąd, niezależnie od intencji ich twórców. Wszak nie istnieje coś takiego jak neutralna architektura selekcji, która nie faworyzowałaby określonych wyników.

Istnieją tylko architektury, które rozkładają ryzyko mniej lub bardziej jawnie, co czyni proces ich projektowania aktem głęboko politycznym. Architektura selekcji, czyli modelowanie AI jako systemu aktywnie filtrującego sygnały i priorytety, sprawia, że każda decyzja projektowa kształtuje przyszłe procesy decyzyjne. Zarządzanie zasobami (Resource Management) oraz jakość usług (Quality of Service), czyli dynamiczna administracja czasem i mocą obliczeniową sensora, stanowią być może najbardziej niedoceniony rdzeń całej problematyki. Dla laika są to terminy brzmiące technicznie, niemal sucho, kojarzące się raczej z instrukcją obsługi routera niż z filozofią. Dla eksperta powinny one jednak brzmieć jak pytanie o samą konstytucję systemu i jego hierarchię wartości. Gdy radar musi w milisekundach decydować o priorytetach, nie wykonuje on wyłącznie jałowego obliczenia.

Ustanawia on bowiem konkretny porządek pierwszeństw, który odzwierciedla priorytety strategiczne danej organizacji w sposób bezlitosny. Użyteczność nie jest tu zwykłą skalą pomocniczą, lecz językiem, w którym instytucja zapisuje to, co uważa za najcenniejsze w swojej misji. Artykuł Fundacji Dobre Państwo o architekturze innowacji słusznie podkreśla, że projekty AI muszą być zakotwiczone w realnej strategii organizacji. Należy je chronić przed tak zwanym teatrem KPI, czyli fetyszyzacją wskaźników, które są całkowicie oderwane od rzeczywistego celu działania. To samo można powiedzieć o każdym sensorze czy systemie wieloagentowym, który bez jasnego celu staje się jedynie kosztownym gadżetem. Nie wystarczy bowiem optymalizować procesów dla samej idei wydajności, jeśli nie wiemy, jaki porządek wartości został wpisany do funkcji celu.

W tym miejscu ekonomia, prawo i informatyka spotykają się w sposób absolutny i nierozerwalny, tworząc nową dyscyplinę zarządzania ryzykiem. Z ekonomii czerpiemy problem niedoboru i użyteczności krańcowej, z prawa zaś kwestie odpowiedzialności, rozliczalności oraz dopuszczalnego poziomu ryzyka. Z informatyki pochodzą mechanizmy reprezentacji, predykcji i optymalizacji, które nadają tym abstrakcyjnym pojęciom formę działającego kodu. Nad nimi wszystkimi unosi się

jednak antropologia polityczna, stawiająca pytania o to, kto ma prawo ustawić ostateczną funkcję celu. Kto decyduje, jaki odsetek błędów będzie tolerowany w imię wyższej konieczności, a kto poniesie konsekwencje za tragiczne w skutkach przeoczenie. To nie są pytania dodatkowe, które można zadać po zakończeniu prac projektowych, lecz samo serce każdego nowoczesnego systemu.

Traktowanie technologii jako zjawiska przyrodniczego jest błędem poznawczym, gdyż rozwój narzędzi cyfrowych wynika z konkretnych decyzji inwestycyjnych i regulacyjnych. Ta intuicja doskonale pasuje do systemów obronnych, gdzie nie ma żadnego technicznego fatum zmuszającego nas do budowania rozwiązań odbierających ludziom sprawstwo. To zawsze jest wybór organizacyjny, toteż tak kluczowa pozostaje kwestia wyjaśnialności, która pozwala człowiekowi zachować kontrolę nad maszyną. Tendencja do ulegania automatyzacji (Automation Bias), czyli skłonność operatorów do bezkrytycznego ufania algorytmom nawet przy ich ewidentnej błędności, stanowi realne zagrożenie dla etyki działań. NIST w swoim raporcie o czterech zasadach xAI podkreśla, że system powinien dostarczać wyjaśnień znaczących dla użytkownika i poprawnych merytorycznie. Ta formuła jest znakomita właśnie dlatego, że nie popada w tani triumfalizm i nie obiecuje nam cudów.

Nie obiecuje ona bowiem całkowitego otwarcia czarnej skrzynki, lecz wyznacza rygor minimalny dla systemu, który ma pozostać używalny w świecie ludzi. W środowisku operacyjnym, zwłaszcza pod presją czasu, nikt nie potrzebuje wykładu filozoficznego do każdego wyniku wygenerowanego przez algorytm. Potrzeba natomiast takiego rodzaju interpretacji, który umożliwi człowiekowi realne zakwestionowanie rekomendacji lub jej w pełni świadome przyjęcie. Znacząca kontrola ludzka (Meaningful Human Control), czyli standard nadzoru zapewniający realną możliwość interwencji, jest tu warunkiem koniecznym, by uniknąć degradacji operatora. xAI nie ma służyć temu, aby model wyglądał uprzejmiej w oczach opinii publicznej czy komisji etycznych. Ma służyć temu, aby człowiek nie został zredukowany do roli biologicznego potwierdzenia dla procesu, którego już dawno nie kontroluje.

W przeciwnym razie człowiek w pętli (human in the loop) staje się liturgią odpowiedzialności bez odpowiedzialności, co jest scenariuszem wysoce niebezpiecznym dla państwa. Stąd bierze się filozoficzna i prawna siła tezy, że odpowiedzialność za czyny należy wyłącznie do ludzi, a nie do maszyn. AI Act nie tworzy podmiotowości prawnej systemów, lecz nakłada konkretne obowiązki na dostawców, wdrażających oraz podmioty eksploatujące te narzędzia. Szczególny nacisk kładzie on na systemy wysokiego ryzyka, wymagając od nich przejrzystości, nadzoru oraz pełnej zgodności z normami. Harmonogram wdrażania tych przepisów jest rozpisany etapami, co dowodzi, że Unia Europejska traktuje AI jako obszar poważnego zarządzania ustrojowego. Zakazane praktyki oraz

obowiązki dotyczące kompetencji w zakresie sztucznej inteligencji zaczęły obowiązywać już od lutego 2025 roku.

Obowiązki wobec modeli ogólnego przeznaczenia wejdą w życie w sierpniu 2025 roku, a pełna stosowalność aktu przypadnie na sierpień 2026 roku. Przewidziano przy tym dłuższe terminy dla części systemów wysokiego ryzyka, które są osadzone w produktach regulowanych innymi przepisami. Jest to istotne, gdyż pokazuje, że współczesna recepcja technologii przeszła od fazy naiwnego zachwytu do fazy dojrzałej, infrastrukturalnej odpowiedzialności. Komputer, jak każdy współczesny biurokrata, bywa szczególnie groźny wtedy, gdy myli pewność z kompetencją, co musimy stale monitorować. Ostatecznie to my decydujemy, czy AI będzie naszym wsparciem, czy też nowym rodzajem cyfrowego determinizmu, któremu bezwolnie się poddamy. W tym wielkim projekcie technologicznym stawką jest bowiem nic innego jak zachowanie ludzkiej podmiotowości w świecie maszyn.

## Architektura odpowiedzialności w cieniu automatyzmu

Nie wolno ulegać złudzeniu, że sama litera prawa, choćby najbardziej rygorystyczna, jest w stanie samodzielnie udźwignąć ciężar etyczny nowych technologii. Prawo potrafi wprawdzie wymusić drobiazgową dokumentację, cykliczne testy czy skomplikowane procedury nadzoru, lecz nie posiada mocy sprawczej, by zadekretować uczciwość tam, gdzie organizacja jej nie pożąda. To właśnie tutaj przebiega najbardziej drażliwa granica między biurokratyczną poprawnością a realnym sprawstwem człowieka. Możemy przecież zaprojektować system, w którym operator formalnie pozostaje w pętli decyzyjnej, a w rzeczywistości zostaje z niej brutalnie wypchnięty przez tempo procesów i narzuconą złożoność. W takim układzie człowiek przestaje być autorem decyzji, stając się jedynie jej ceremonialnym, niemal liturgicznym podpisem pod werdyktem maszyny. Wszak human in the loop staje się wtedy liturgią odpowiedzialności bez odpowiedzialności, gdzie forma przysłania całkowity zanik ludzkiej podmiotowości.

Zjawisko to, znane jako automation bias, czyli psychologiczna tendencja do bezkrytycznego ufania wynikom generowanym przez algorytmy, nie jest jedynie błędem jednostki, lecz efektem architektury sytuacji. Im bardziej system działa szybko, celnie i przekonująco, tym silniejsza staje się pokusa, by przestać kwestionować jego wskazania. Wtedy odpowiedzialność zachowuje swoją zewnętrzną, ludzką formę, ale traci całą treść realnego działania. Tego właśnie powinniśmy obawiać się najbardziej w nadchodzącej dekadzie. Nie buntu maszyn, który pozostaje domeną kina klasy B, lecz społecznej i instytucjonalnej zgody na pozór kontroli. Współczesna nauka o systemach

inteligentnych prowadzi nas bowiem do wniosku paradoksalnego: im bardziej technologia rośnie w siłę, tym mniej wystarcza nam ona sama.

Potrzeba nam dzisiaj znacznie więcej niż tylko sprawniejszych procesorów; potrzeba więcej prawa, rzetelnego audytu, głębokiej filozofii odpowiedzialności oraz organizacyjnej samodyscypliny. Musimy monitorować modele długo po ich wdrożeniu, ponieważ środowisko operacyjne zmienia się szybciej niż zestawy danych treningowych. Trzeba badać granice wiedzy systemu, gdyż nowość w świecie rzeczywistym nigdy nie pyta o laboratoryjne benchmarki. Niezbędne jest dbanie o wyjaśnialność adekwatną do kompetencji użytkownika, bo bez niej człowiek staje się jedynie zakładnikiem interfejsu. Komputer, jak każdy współczesny biurokrata, bywa szczególnie groźny wtedy, gdy myli pewność z kompetencją, narzucając nam swoją logikę jako jedyną możliwą prawdę.

Trzeba uparcie zakotwiczać projekty AI w realnych celach strategicznych, ponieważ w przeciwnym razie wskaźniki KPI zaczną zastępować nam rzeczywistość. Równie ważne jest konsekwentne odmawianie maszynom statusu moralnego, gdyż inaczej instytucje zbyt łatwo użyją ich jako wygodnej zasłony dymnej dla własnych, często błędnych decyzji. To nie jest wyraz konserwatywnej ostrożności czy lęku przed nowym, lecz jedyna poważna i dojrzała nowoczesność. Możemy więc już teraz sformułować domknięcie naszych rozważań. Klasyfikacja sygnałów, planowanie misji, ELINT, czyli wywiad elektroniczny oparty na przechwytywaniu sygnałów, oraz techniki Track before Detect nie są rozproszonymi rozdziałami technicznej encyklopedii.

Wszystkie te moduły, wraz z rozpoznaniem nuklearnym, zarządzaniem zasobami czy obliczeniami kwantowymi, tworzą wspólną teorię nowoczesnego działania w warunkach głębokiej niepewności. Każdy z nich odpowiada na inne, fundamentalne pytanie o naszą relację ze światem. Co właściwie widzimy w szumie danych? Jak długo zdołamy utrzymać stabilny obraz sytuacji? W jaki sposób rozdzielimy istotne ślady od tła? Jak przydzielimy ograniczoną uwagę tam, gdzie zagrożeń jest zbyt wiele? Jak wreszcie nie pomylimy radykalnej nowości z tym, co już znamy i potrafimy sklasyfikować?

Nad tymi wszystkimi kwestiami unosi się jednak pytanie najstarsze i najważniejsze: kto za to wszystko odpowiada? Odpowiedź źródłowa, której należy bronić z całą mocą, brzmi jasno i bezalternatywnie: odpowiada człowiek. Zawsze człowiek. System może nas wspierać, ostrzegać, optymalizować procesy i wyjaśniać swoje kroki, ale nie może ponosić winy, troski ani zobowiązania. Nie może być depozytariuszem odpowiedzialności, ponieważ nie jest kimś, lecz czymś. I właśnie dlatego powinien być projektowany tak, aby nigdy nie pozwolić nam o tej fundamentalnej różnicy zapomnieć.

To domknięcie nie jest wygodne dla współczesnego menedżera czy polityka. Atakuje ono konformizm, który chciałby uwierzyć, że wystarczy więcej danych i ładniejszy dashboard, aby zniknęły stare problemy władzy, błędu i sumienia. One nie znikną, a wręcz przeciwnie – staną się bardziej wyrafinowane i trudniejsze do uchwycenia. AI nie usuwa polityki, lecz ją koduje w swoich wagach i progach decyzyjnych. Nie usuwa etyki, lecz brutalnie ujawnia, czy w ogóle ją jako wspólnota posiadamy. AI nie usuwa odpowiedzialności, a jedynie wystawia na próbę naszą odwagę, by nie chować się za ekranem monitora.

Najdojrzalsza obrona tez przedstawionych w tym artykule nie polega na bezkrytycznym zachwycie nad technologią. Polega ona na żądaniu, by każda przewaga poznawcza maszyny była osadzona w przewadze instytucjonalnej rozumu nad wygodą automatyzmu. Jeśli tej przewagi zabraknie, wszystkie nasze inteligentne systemy okażą się ostatecznie tylko bardzo szybkimi maszynami do produkowania dobrze ubranej bezradności. Nie stoimy bowiem wobec pojedynczej rewolucji technicznej, lecz wobec rekonfiguracji całego porządku poznawczego nowoczesnych instytucji. Wszystkie omawiane technologie są kolejnymi wariantami tej samej odpowiedzi na pytanie: jak działać rozumnie tam, gdzie świat dostarcza zbyt wielu danych i zbyt mało czasu.

Sztuczna inteligencja nie zastępuje ludzkiego rozumu, lecz brutalnie obnaża punkty, w których rozum instytucjonalny był dotąd prowizoryczny lub zwyczajnie niewydolny. Pokazuje, że nowoczesna władza nie polega już na samym posiadaniu informacji, lecz na zdolności ich filtrowania i przekuwania w decyzję szybciej niż chaos. Fundacja Dobre Państwo, analizując systemowe podejście do innowacji, trafnie wskazuje na potrzebę organizacyjnego szkieletu, który wiąże technologię ze strategią. Bez tego wiązania projekty AI będą jedynie dryfować w stronę kosztownego chaosu, marnując potencjał i zaufanie społeczne.

Odpowiedzialność pozostaje ludzka nie z sentymentu do tradycji, ale dlatego, że tylko człowiek może być nośnikiem obowiązku i rozliczalności. Maszyna może kalkulować i szeregować wyniki, a nawet tłumaczyć własne procesy w zakresie, na jaki pozwala jej architektura wyjaśnialności. Nie może jednak stanąć wobec normy prawnej czy moralnej tak, jak staje wobec niej osoba. Nie może ponieść odpowiedzialności wobec ofiary błędu ani wobec wspólnoty politycznej. Właśnie dlatego europejskie ramy regulacyjne, w tym AI Act, nie nadają systemom żadnej quasi-podmiotowości, lecz nakładają konkretne obowiązki na ich dostawców i użytkowników.

Warto pamiętać o kalendarzu tych zmian, gdyż AI Act wszedł w życie 1 sierpnia 2024 roku, a jego kolejne etapy sukcesywnie zmieniają krajobraz prawny. Obowiązki dotyczące zakazanych praktyk zaczęły obowiązywać od 2 lutego 2025 roku, a modele ogólnego przeznaczenia zostaną objęte regulacjami od sierpnia 2025 roku. Pełna stosowalność aktu, przypadająca na 2 sierpnia 2026 roku, to moment, w którym Meaningful Human Control, czyli standard realnego nadzoru nad

systemem, przestanie być tylko postulatem etycznym. Stanie się on twardym wymogiem dla systemów wysokiego ryzyka, wymuszając na organizacjach odejście od rytualnych podpisów na rzecz faktycznej kontroli.

Spór o AI nie jest już sporem o to, czy modele będą lepsze – to jest niemal przesądzone przez postęp techniczny. Spór dotyczy tego, czy rozwój tych modeli zostanie trwale związany z rozwojem architektury odpowiedzialności. Jeżeli ten proces się nie powiedzie, otrzymamy systemy imponujące poznawczo, ale infantylne ustrojowo. Będą one wsysały człowieka w pozór kontroli, gdzie human in the loop pozostanie jedynie w dokumentacji technicznej. Wtedy operator będzie podpisywał werdykty, których nie rozstrzyga, i odpowiadał za skutki, nad których przyczynami nie panuje.

Właśnie temu ma przeciwdziałać xAI, czyli technologia wyjaśnialnej sztucznej inteligencji, która nie powinna być traktowana jako kosmetyka zaufania. Jej rolą jest przywrócenie człowiekowi prawa do sensownego sprzeciwu wobec rekomendacji maszyny. NIST w swoich czterech zasadach wyjaśnialnej AI podkreśla, że system musi dostarczać uzasadnień znaczących dla użytkownika i poprawnych względem procesu generowania wyniku. To ujęcie nie obiecuje nam metafizycznego wglądu w duszę sieci neuronowej, ale żąda czegoś znacznie ważniejszego. Żąda, by człowiek otrzymał taki rodzaj argumentacji, który pozwoli mu na rzetelną ocenę ryzyka i ewentualne odrzucenie sugestii algorytmu.

W systemach sensorowych i obronnych to rozróżnienie między zrozumieniem a posłuszeństwem jest decydujące dla bezpieczeństwa misji. Techniki takie jak heatmapy czy lokalne aproksymacje nie służą temu, by model wyglądał na bardziej przyjazny dla użytkownika. Mają one sprawić, by operator nie stał się wyłącznie notariuszem wyniku wyprodukowanego przez czarną skrzynkę. Jeśli wyjaśnienie nie zwiększa realnej zdolności do zakwestionowania rekomendacji, jest jedynie nowocześniejszą i lepiej podświetloną odmianą bezkrytycznego posłuszeństwa. Ciężar tego problemu staje się szczególnie widoczny w zarządzaniu zasobami i jakością usług (QoS).

Zarządzanie zasobami, czyli dynamiczna administracja czasem i mocą obliczeniową sensora, może brzmieć jak suchy temat inżynieryjny, ale w istocie jest to teoria sprawiedliwego rozdziału uwagi. Gdy radar musi zdecydować, czy poświęcić milisekundę na śledzenie celu, czy na walkę elektroniczną, działa w logice mennicy aksjologicznej. Ktoś musiał wcześniej zdecydować, jak ważyć błąd śledzenia wobec ryzyka utraty obiektu. To jest właśnie aksjologia progu alarmowego – uznanie, że wybór czułości sensora jest głęboką decyzją o wartościach, a nie tylko parametrem technicznym. System może optymalizować procesy perfekcyjnie, a jednak robić to dla celów, które zostały źle zdefiniowane przez człowieka.

W tym kontekście powraca wątek inspiracji kwantowych, takich jak model Isinga czy obliczenia adiabatyczne, które pokazują nową odwagę współczesnej nauki. Ich siła nie polega na futurystycznym połysku, lecz na pytaniu, czy sama fizyka nie może przejąć części ciężaru obliczeniowego związanego z redukcją złożoności. Redukcja ta, będąca procesem brutalnego upraszczania rzeczywistości przez AI, jest niezbędna do działania w czasie rzeczywistym, ale zawsze odbywa się kosztem pominięcia pewnych aspektów świata. Dojrzałość instytucji polega na świadomości tych kosztów i budowaniu systemów, które potrafią działać bez wydziałowych poręczy, łącząc fizykę z etyką i prawem.

Prawdziwa biegłość nie kończy się na zakupie i wdrożeniu modelu, lecz na zdolności stworzenia pełnego cyklu życia odpowiedzialności wokół niego. Od audytu danych treningowych, przez certyfikację, aż po ciągły monitoring po wdrożeniu, który NIST uznaje za kluczowy element bezpieczeństwa. W systemach wysokiego ryzyka benchmark jest tylko chwilową fotografią, podczas gdy realna operacja to zmienny i nieprzewidywalny klimat. Klimat ten ma tę nieprzyjemną właściwość, że nie pyta algorytmu o zdanie, zanim postawi go przed zupełnie nowym wyzwaniem.

Nadchodzi świat, w którym instytucje będą albo umiały podporządkować moc obliczeniową porządkowi odpowiedzialności, albo same staną się służebnicami własnych systemów selekcji. Można budować AI jako infrastrukturę rozumu publicznego, związaną z prawem i realnym nadzorem, albo jako infrastrukturę wygodnego alibi. Ta druga droga jest prostsza i politycznie kusząca, bo pozwala twierdzić, że każda decyzja była racjonalna, skoro została policzona przez maszynę. Jednak to właśnie na tej drodze najczęściej dochodzi do sytuacji, w której inteligentny system produkuje bardzo stare i dobrze nam znane nieszczęście.

Najkrótsza puenta brzmi więc następująco: klasyfikacja sygnałów, ELINT, obliczenia kwantowe i etyka tworzą razem spójną teorię odpowiedzialnego działania pod niepewnością. Kto to rozumie, ten nie pyta już naiwnie, czy sztuczna inteligencja będzie kiedyś moralna. Pyta raczej, czy my sami pozostaniemy wystarczająco moralni, aby nie użyć jej jako zasłony dla własnej nieodpowiedzialności. I to jest pytanie naprawdę godne dojrzałego wieku maszyn, w którym technologia staje się ostatecznym testem dla naszego człowieczeństwa.

> Sztuczna inteligencja nie usuwa niepewności z pola walki, lecz jedynie zmienia cenę, jaką płacimy za jej ignorowanie. Prawdziwy test dla nowoczesnych instytucji nie polega na tym, jak precyzyjnie maszyny potrafią liczyć, lecz na tym, czy człowiek zachowa odwagę, by w decydującym momencie powiedzieć maszynie 'nie'. W świecie wiecznej zmienności, nasza podmiotowość jest jedynym zasobem, którego nie da się zoptymalizować ani poddać outsourcingowi.

## Inspiracje

[1] 


**"Ancient towns of China"** Wang Baochuan, Zhang Yuanijing

*Palgrave Macmillan | ISBN: 9789819587452*

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: a16428ba-ec9c-417b-ba9c-e47411114a66