Algorytmiczny feudalizm czy cyfrowa republika?

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Artykuł analizuje ewolucję organizacji w dobie sztucznej inteligencji, wskazując na konieczność przejścia od fascynacji techniczną sprawnością modeli do głębokiej odpowiedzialności cywilizacyjnej. Autor stawia tezę, że dojrzała instytucja AI nie powinna traktować algorytmów jedynie jako narzędzi optymalizacji, lecz jako jednostki decyzyjne wymagające sformalizowanych ram etycznych – swoistej konstytucji. Tekst definiuje filary odpowiedzialnego wdrażania technologii, kładąc nacisk na inżynierię odpowiedzialności, myślenie ex ante oraz rzetelną analizę ryzyka. Zamiast testować rozwiązania na społeczeństwie, organizacje muszą budować systemy z uwzględnieniem ich wpływu na grupy wrażliwe i podmiotowość człowieka. Wybór między algorytmicznym feudalizmem a cyfrową republiką zależy od integracji wyjaśnialności i moralnego kompasu z procesami technologicznymi.

This article analyzes the evolution of organizations in the age of artificial intelligence, pointing to the need to shift from a fascination with the technical efficiency of models to a profound civilizational responsibility. The author argues that a mature AI institution should not treat algorithms merely as optimization tools, but as decision-making entities requiring a formalized ethical framework—a kind of constitution. The text defines the pillars of responsible technology implementation, emphasizing responsibility engineering, ex ante thinking, and reliable risk analysis. Instead of testing solutions on society, organizations must build systems that consider their impact on vulnerable groups and human agency. The choice between algorithmic feudalism and a digital republic hinges on integrating explainability and a moral compass with technological processes.

Współczesne instytucje wdrażające sztuczną inteligencję stają przed ontologicznym wyzwaniem: przejściem od traktowania technologii jako narzędzia optymalizacji do uznania jej za podmiot wymagający cywilizacyjnej odpowiedzialności. Niniejszy tekst

dekonstruuje mit AI jako neutralnego rozwiązania, wskazując, że bez rygorystycznych fundamentów – takich jak audytowalność, przejrzystość algorytmiczna i realny nadzór ludzki – organizacje ryzykują budowę systemów, które zamiast służyć wspólnocie, tworzą mechanizmy wykluczenia i proceduralnej przemocy. Autorzy argumentują, że dojrzałość instytucjonalna objawia się w zdolności do projektowania systemów z uwzględnieniem skutków ex ante, rzetelnym zarządzaniu danymi oraz świadomej rezygnacji z automatyzacji tam, gdzie ryzyko naruszenia godności ludzkiej przewyższa korzyści z efektywności. W dobie cyfrowego feudalizmu, gdzie algorytmy mogą niepostrzeżenie determinować szanse życiowe obywateli, kluczową kompetencją strategiczną staje się kultura sceptycyzmu i prawo do odmowy. Tekst stanowi diagnozę konieczności powiązania inżynierii z etyką i prawem, postulując, że przyszłość odpowiedzialnej AI zależy od ustanowienia instytucji zdolnych do sprawowania realnej kontroli nad kodem, który kształtuje naszą rzeczywistość społeczno-polityczną.

SŁOWA KLUCZOWE

algorytmiczny feudalizm | cyfrowa republika | human-in-the-loop | wyjaśnialność | audyt ciągły | red teaming | agentic AI | sandboxing | fairness | retencja danych | inżynieria odpowiedzialności | zarządzanie wiedzą | odpowiedzialność środowiskowa | mapa interesariuszy | ontologia

Od technicznej sprawności do cywilizacyjnej odpowiedzialności

Dojrzała instytucja AI nie zaczyna swojej drogi od dylematu, który model językowy kupić, jak go zintegrować z istniejącym ekosystemem czy w jaki sposób dotrenować go na własnych zasobach. Prawdziwa dojrzałość objawia się w postawieniu fundamentalnych pytań o naturę władzy, jaką zamierzamy powierzyć systemowi, oraz o to, nad kim ta władza będzie w istocie sprawowana. Musimy precyzyjnie określić, kto będzie weryfikował te decyzje, kto wyznaczy im nieprzekraczalne granice i – co najważniejsze – kto weźmie pełną odpowiedzialność za ich nieuniknione skutki. To przesunięcie akcentu z technicznej sprawności modelu na cywilizacyjną odpowiedzialność za jego działanie jest zmianą o charakterze ontologicznym. Bez tej refleksji technologia staje się jedynie

kolejnym narzędziem bezładnej ekspansji, pozbawionym kompasu moralnego i prawnego, co w dłuższej perspektywie musi prowadzić do instytucjonalnego chaosu.

Organizacje niedojrzałe wciąż postrzegają sztuczną inteligencję wyłącznie przez pryzmat narzędzia produktywności, mającego optymalizować arkusze kalkulacyjne i przyspieszać rutynowe zadania biurowe. Tymczasem dojrzała instytucja AI traktuje te systemy jako pełnoprawne jednostki decyzyjne, które wymagają własnej konstytucji, przejrzystego archiwum, niezależnego sądu oraz sformalizowanych procedur odwoławczych. Taka struktura musi być wsparta kulturą bezwzględnej prawdomówności i stałego nadzoru, co może brzmieć dla technokratów zbyt patetycznie lub wręcz anachronicznie. Warto jednak pamiętać, że algorytm klasyfikujący pacjenta, kandydata do pracy czy kredytobiorcę nie jest jedynie niewinną, matematyczną kalkulacją wykonaną w próżni. Jest on brutalną interwencją w cudzy świat możliwości, kształtującą życiorysy i determinującą dostęp do dóbr podstawowych, dlatego wymaga rygoru równego instytucjom państwowym.

Kardynalnym błędem współczesnych organizacji jest traktowanie odpowiedzialnej AI jako powierzchownej warstwy reputacyjnej, którą nakłada się na gotowy produkt niczym makijaż przed sesją zdjęciową. Najpierw budujemy skomplikowany system, a dopiero później, pod presją działu PR, dopisujemy do niego ogólne zasady etyczne, które mają uspokoić sumienia zarządu i akcjonariuszy. Najpierw wdrażamy pełną automatyzację procesów, a potem gorączkowo szukamy argumentów potwierdzających, że nasze działania nie naruszają elementarnych norm społecznych. Najpierw optymalizujemy koszty operacyjne, a dopiero w obliczu kryzysu zamawiamy zewnętrzny audyt, który ma zalegalizować dokonane już fakty i uspokoić regulatora. Taki porządek jest tragicznie odwrócony, przypominając budowę wieżowca bez fundamentów, w którym stabilność konstrukcji sprawdza się dopiero po wprowadzeniu wszystkich lokatorów.

Odpowiedzialna AI wymaga myślenia ex ante, czyli projektowania systemów z uwzględnieniem skutków przed ich wystąpieniem, co nie powinno być postrzegane jako biurokratyczna przeszkoda. Taka inteligencja projektowa jest w istocie warunkiem koniecznym bezpieczeństwa, podobnie jak ma to miejsce w inżynierii lotniczej czy nowoczesnej medycynie. W lotnictwie nikt nie zaczyna przecież od lotu z pasażerami, aby dopiero wysoko w chmurach sprawdzać, czy konstrukcja skrzydeł była właściwym pomysłem inżynieryjnym. W AI niestety wciąż zbyt często testuje się skrzydła na społeczeństwie, traktując obywateli jak darmowe króliki doświadczalne w wielkim eksperymencie technologicznym. Dojrzała instytucja musi najpierw precyzyjnie określić domenę ryzyka, rozumiejąc, że inaczej projektuje się system do streszczania notatek, a zupełnie inaczej asystenta onkologa.

Kategorie ryzyka nie są pojęciami abstrakcyjnymi, lecz wynikają bezpośrednio z rodzaju i skali potencjalnej szkody, jaką system może wyrządzić konkretnej jednostce lub całej grupie. Musimy

pytać, czy błąd algorytmu może pozbawić kogoś prawa do zasiłku, szansy na zatrudnienie, zdolności kredytowej, właściwego leczenia czy wreszcie dobrego imienia. Należy rozważyć, czy system dotyczy grup szczególnie wrażliwych, takich jak dzieci, pacjenci czy mniejszości, które mają ograniczoną możliwość obrony swoich interesów przed cyfrowym gigantem. Kluczowe staje się ustalenie, czy decyzja podjęta przez maszynę jest w ogóle odwracalna i czy człowiek posiada realną ścieżkę odwoławczą w świecie zdominowanym przez kod. Czy skutki awarii będą miały charakter jednostkowy, czy może przerodzą się w błąd systemowy, infekujący całą tkankę społeczną i podważający zaufanie do państwa?

Pytania te nie stanowią jedynie etycznego dodatku do inżynierii, lecz są samą istotą inżynierii odpowiedzialności, bez której technologia staje się niebezpiecznym narzędziem chaosu. Drugim filarem tej konstrukcji jest rzetelna mapa interesariuszy, która w wielu projektach AI upada moralnie już na etapie wstępnego planowania i definiowania celów. Zbyt często za interesariusza uznaje się wyłącznie klienta biznesowego, właściciela procesu oraz dostawcę technologii, całkowicie ignorując tych, których system dotyczy bezpośrednio. Tymczasem prawdziwymi interesariuszami są osoby oceniane przez algorytmy, pracownicy operujący na ich rekomendacjach, a także regulatorzy i całe społeczeństwo obywatelskie. Jeśli system wpływa na życie obywatela, to nie jest on jedynie technicznym użytkownikiem końcowym, lecz podmiotem prawa, którego godność musi być bezwzględnie chroniona.

Trzecim filarem dojrzałości jest zarządzanie danymi, przy czym zaufaną AI należy budować na fundamentach znacznie solidniejszych niż te opisane w kolorowych broszurach marketingowych. Zaufane dane to nie tylko zbiory kompletne i aktualne, ale przede wszystkim dane legalne, proporcjonalne, reprezentatywne i wolne od ukrytej dyskryminacji. Organizacja musi posiadać pełną wiedzę o historii danych, ich ograniczeniach, lukach oraz specyficznym kontekście, w jakim zostały pierwotnie wytworzone i przetworzone. Dane pozbawione kontekstu są jak zeznania świadka wyrwane z długiego procesu sądowego: mogą być formalnie prawdziwe, a jednak prowadzić do głęboko niesprawiedliwego wyroku. Bez tej świadomości ryzykujemy wpadnięcie w feudalizm algorytmiczny, czyli model władzy, w którym dane organizują hierarchię społeczną bez realnej odpowiedzialności ich dysponentów.

Czwartym filarem jest retencja, rozumiana jako rozsądna pamięć instytucjonalna, która pozwala na odtworzenie ścieżki decyzyjnej w razie wystąpienia błędu lub dryfu modelu. W epoce wszechobecnej inwigilacji trzeba przechowywać wystarczająco dużo informacji, aby móc zbadać incydent, ale nie wolno gromadzić wszystkiego pod pretekstem mglistej, przyszłej użyteczności. Wieczna retencja jest niebezpieczną pokusą biurokratycznego smoka, który gromadzi skarby danych, nie potrafiąc ich zrozumieć, a jedynie zwiększając ryzyko wycieków i nadużyć. Dojrzała

instytucja doskonale zna różnicę między pamięcią służącą odpowiedzialności a zachłannością, która jest jedynie narzędziem totalnej kontroli nad jednostką. Pamięć powinna być filtrem, który zatrzymuje to, co niezbędne do sprawiedliwości, odrzucając jednocześnie balast zbędnych informacji naruszających prywatność i autonomię człowieka.

Piątym filarem jest architektura wyjaśnialności, która nie polega na podaniu jednej, uniwersalnej odpowiedzi na pytanie o przyczyny konkretnej decyzji podjętej przez system. Chodzi tu raczej o wielowarstwowy system uzasadnień, dostosowany do potrzeb różnych odbiorców: od zwykłego użytkownika po wyspecjalizowanego audytora i państwowego regulatora. Użytkownik powinien otrzymać wyjaśnienie praktyczne i zrozumiałe, podczas gdy profesjonalista musi mieć wgląd w metryki, wersje systemu oraz szczegółowe ścieżki decyzyjne. Wyjaśnialność pełni funkcję tłumacza między światami: prawnym, społecznym, biznesowym i matematycznym, zapobiegając sytuacji, w której AI staje się wieżą Babel predykcji. Bez tego mechanizmu systemy stają się czarnymi skrzynkami, których działanie jest równie nieprzeniknione, co wyroki dawnych bóstw, co wyklucza jakąkolwiek formę demokratycznej kontroli.

Szóstym filarem jest realny human-in-the-loop, a nie jego teatralna atrapa, która ma jedynie uspokajać opinię publiczną i zdejmować ciężar z barków projektantów. Człowiek musi znajdować się w pętli decyzyjnej wszędzie tam, gdzie jego osąd jest moralnie i prawnie konieczny, posiadając przy tym realne warunki do sprawowania nadzoru. Oznacza to konieczność zapewnienia mu odpowiednich kompetencji, czasu, dostępu do pełnej informacji oraz – co najważniejsze – prawa do zakwestionowania rekomendacji maszyny. Człowiek w pętli bez władzy jest tylko dekoracją sumienia, figurantem, który swoim podpisem legitymizuje procesy, nad którymi w rzeczywistości nie ma żadnej kontroli. Jeśli urzędnik korzysta z algorytmicznej oceny ryzyka, musi mieć możliwość jej odrzucenia bez obawy o sankcje ze strony przełożonych nastawionych na ślepą efektywność.

Siódmym filarem jest audyt ciągły, ponieważ statyczna weryfikacja przedwdrożeniowa, choć wartościowa, nie wystarcza w dynamicznie zmieniającym się środowisku technologicznym i społecznym. Modele AI żyją, ewoluują i reagują na nowe dane, zachowania użytkowników oraz zmieniające się regulacje prawne, co sprawia, że system bezpieczny wczoraj może być wadliwy dzisiaj. Dlatego audyt musi obejmować stałe monitorowanie produkcyjne, analizę skarg, kontrolę halucynacji oraz badanie różnic w traktowaniu poszczególnych grup społecznych. Dojrzała instytucja nie pyta retorycznie, czy system był bezpieczny w dniu uroczystej prezentacji przed zarządem, lecz czy jest on bezpieczny w tej konkretnej minucie. Wymaga to gotowości do natychmiastowego wycofania narzędzia z eksploatacji, jeśli tylko zostanie wykryty niebezpieczny dryf lub naruszenie fundamentalnych zasad etycznych.

Ósmym filarem jest red teaming, czyli praktyka polegająca na konfrontowaniu systemu z inteligentnym przeciwnikiem, który stara się celowo wykazać jego słabości i podatności. Nie robimy tego z braku zaufania do inżynierów, lecz z głębokiego przekonania, że rzeczywistość bywa złośliwa, twórcza i rzadko bywa uprzejma dla twórców technologii. Red team powinien testować obejścia zabezpieczeń, próby manipulacji kontekstem oraz podatność modelu na stereotypy czy błędne instrukcje, które mogą prowadzić do katastrofalnych skutków. System, który działa poprawnie tylko wtedy, gdy wszyscy użytkownicy są mili, kompetentni i przewidywalni, nie jest w żadnym razie gotowy na spotkanie z realnym światem. Taki produkt jest gotowy jedynie na efektowne demo, które ma uwieść inwestorów, ale nie nadaje się do pełnienia odpowiedzialnych funkcji społecznych.

Fundamenty dojrzałości: jak budować odpowiedzialną infrastrukturę AI

Dziewiątym filarem, bez którego gmach odpowiedzialnej technologii osunąłby się w niekontrolowany chaos, jest sandboxing oraz rygorystyczne stopniowanie autonomii. Agentic AI, czyli systemy posiadające zdolność do samodzielnego podejmowania działań i operowania w imieniu użytkownika, nie mogą być wpuszczane do pełnego środowiska produkcyjnego bez kwarantanny. Autonomia powinna być przyznawana warstwowo, niczym kolejne stopnie wtajemniczenia: od prostych rekomendacji, przez ograniczone operacje pod nadzorem, aż po działania automatyczne wyłącznie tam, gdzie ryzyko jest rozpoznane i w pełni odwracalne. System musi posiadać twarde limity uprawnień, blokady kosztowe oraz mechanizmy awaryjnego zatrzymania, które działają niezależnie od logiki samego agenta. Autonomiczny agent pozbawiony wyraźnych granic przypomina bowiem stażystę, któremu przez pomyłkę wręczono klucze do skarbca, serwerowni i profilu premiera w mediach społecznościowych. Entuzjazm takiego pracownika bywa godny podziwu, jednak jego rozsądek rzadko dorównuje skali potencjalnych zniszczeń, jakie może wywołać jedno niefortunne kliknięcie.

Dziesiątym filarem jest fairness, czyli sprawiedliwość, rozumiana nie jako modne, korporacyjne zaklęcie, lecz jako rygorystyczna i powtarzalna procedura operacyjna. Dojrzała instytucja nie ogranicza się do deklaracji, że jej systemy są bezstronne, lecz z niemal inżynierską precyzją pokazuje, w jaki sposób tę bezstronność weryfikuje. Analizuje konkretne grupy społeczne, stosuje mierzalne metryki dyskryminacji i definiuje, jak rozumie krzywdę w specyficznym kontekście danej usługi czy procesu. Wymaga to dokumentowania wyjątków, badania wpływu cech zastępczych na wynik oraz aktywnego konsultowania skutków algorytmicznych decyzji z osobami, których one

bezpośrednio dotyczą. Fairness nie jest bowiem stanem, który raz osiągnięto i można o nim zapomnieć, lecz ciągłą praktyką utrzymywania sprawiedliwości w warunkach zmiennych danych i ewoluujących norm społecznych. To zadanie wymagające biegłości w matematyce, prawie, socjologii i ekonomii, ale przede wszystkim wymagające głębokiej, instytucjonalnej pokory wobec własnych ograniczeń.

Jedenastym filarem jest Knowledge Management, czyli zarządzanie wiedzą, które stanowi kluczowy pomost między ludzkim doświadczeniem a maszynową operacyjnością. Dojrzała sztuczna inteligencja nie powinna być karmiona wyłącznie surowymi danymi, lecz musi czerpać z uporządkowanej wiedzy organizacyjnej, eksperckiej oraz normatywnej. Proces ten wymaga starannej kurateli, translacji oraz dyseminacji wiedzy, aby system nie stał się jedynie generatorem statystycznych prawdopodobieństw. Kuratela oznacza tu selekcję i weryfikację treści, podczas gdy translacja to żmudny przekład zasad etycznych i procedur na konkretne ontologie, reguły i parametry systemowe. Dyseminacja z kolei gwarantuje, że wnioski płynące z działania AI wracają do ludzi, wzbogacając kulturę organizacyjną i zapobiegając degradacji kompetencji pracowników. Bez tego sprzężenia zwrotnego technologia nie uczy organizacji niczego nowego, a jedynie automatyzuje jej pamięć, a co gorsza – czasem automatyzuje także jej zbiorową amnezję.

Dwunastym filarem jest odpowiedzialność środowiskowa, przypominająca nam, że AI nie jest eterycznym duchem czystej wiedzy, lecz bytem o konkretnym ciężarze fizycznym. Każdy model ma swój ślad materiałowy, energetyczny i wodny, a centra danych wymagają gigantycznych nakładów infrastrukturalnych oraz skomplikowanych łańcuchów dostaw. Odpowiedzialna strategia nie może mówić o dobru wspólnym, ignorując jednocześnie ekologiczne koszty generowania kolejnych tokenów i utrzymywania gigantycznych klastrów obliczeniowych. Potrzebujemy energooszczędnych architektur, optymalizacji procesów oraz przede wszystkim rozsądnych decyzji o tym, kiedy potężny model jest naprawdę niezbędny. Nie każda sprawa wymaga przecież użycia cyfrowego młota pneumatycznego, który zużywa energię małego miasteczka, by rozwiązać trywialny problem. Czasem w zupełności wystarczy prosty śrubokręt w postaci klasycznego algorytmu, a czasem po prostu szczera i bezpośrednia rozmowa z drugim człowiekiem.

Trzynastym filarem jest prawo do sprzeciwu i korekty, oparte na założeniu, że system AI w sektorze krytycznym prędzej czy później popełni błąd. Dojrzałość instytucji mierzy się nie tym, że nigdy się nie myli, lecz tym, jak szybko, jawnie i sprawiedliwie potrafi naprawić skutki swoich potknięć. Osoba dotknięta decyzją algorytmiczną musi mieć realną możliwość uzyskania wyjaśnienia, zakwestionowania wyniku oraz kontaktu z kompetentnym pracownikiem. Błąd algorytmiczny, od którego nie ma żadnej drogi odwołania, staje się bowiem współczesną, technokratyczną formą proceduralnej przemocy. Człowiek zostaje w takiej sytuacji zraniony podwójnie: najpierw samą

decyzją, a następnie dojmującą niemożnością rozmowy z kimkolwiek, kto mógłby powiedzieć: „rozumiem, sprawdzimy to”. Etyka bez infrastruktury jest w tym przypadku jedynie kazaniem do serwera, które nie przynosi ulgi ani sprawiedliwości tym, którzy zostali przez system pominięci lub skrzywdzeni.

Czternastym filarem jest kultura organizacyjnego sceptycyzmu, bez której wszystkie spisane standardy i kodeksy pozostaną jedynie martwą literą na papierze. Pracownicy muszą posiadać prawo do głośnego mówienia o ryzyku, kwestionowania założeń projektowych i zatrzymywania wdrożeń bez strachu przed etykietą hamulcowego postępu. Sceptyk w projekcie technologicznym nie jest wrogiem innowacji, lecz pełni funkcję jej układu odpornościowego, chroniąc organizację przed jej własnym, ślepym entuzjazmem. Instytucja, która systemowo wycisza głosy krytyczne, staje się immunologicznie bezbronna wobec własnego sukcesu, co zazwyczaj prowadzi do spektakularnych katastrof. Największe tragedie w historii techniki rzadko zaczynały się od braku inteligencji twórców, a znacznie częściej od kultury, w której mądrzy ludzie przestali mówić prawdę. Przestali, ponieważ prawda psuła optymistyczny harmonogram i nie pasowała do narracji o nieuchronnym triumfie nowoczesności.

Piętnastym filarem jest edukacja użytkowników i profesjonalistów, która wykracza daleko poza proste szkolenie z obsługi nowego interfejsu. Musimy uczyć o ograniczeniach, halucynacjach modeli, ukrytych uprzedzeniach oraz o tym, jak interpretować wyniki w sposób krytyczny i odpowiedzialny. Lekarz, urzędnik czy prawnik potrzebują zupełnie innych kompetencji, by bezpiecznie korzystać z narzędzi AI, niż analityk finansowy czy nauczyciel w szkole podstawowej. Pojedyncza prezentacja o tym, jak pisać skuteczne prompty, nie jest edukacją, lecz przypomina raczej cyfrowe rozdawanie zapałek w składzie pełnym benzyny. Edukacja musi formować nie tylko umiejętność technicznego użycia narzędzia, ale przede wszystkim nawyk zadawania fundamentalnego pytania: czy w tej konkretnej sprawie w ogóle powinienem użyć tego systemu? Prawdziwa wiedza objawia się bowiem nie w biegłości operowania suwakami, lecz w zrozumieniu momentu, w którym należy je puścić.

Szesnastym filarem jest przejrzystość wobec społeczeństwa, szczególnie istotna w przypadku instytucji publicznych, które decydują o prawach i obowiązkach obywateli. Państwo powinno prowadzić jawne rejestry zastosowań AI, publikować rzetelne oceny ryzyka oraz wyjaśniać, gdzie maszyna wspiera decyzję, a gdzie człowiek zachowuje pełną sprawczość. W zdrowej demokracji obywatel nie może dowiadywać się po latach, że był klasyfikowany przez system, o którego istnieniu nie miał najmniejszego pojęcia. Tajna automatyzacja administracji jest fundamentalnie sprzeczna z duchem państwa prawa i niszczy zaufanie do instytucji publicznych. Państwo może oczywiście posiadać swoje sekrety w sprawach bezpieczeństwa narodowego, ale nie powinno mieć ukrytych

maszyn decyzyjnych tam, gdzie rozstrzyga o codziennych sprawach zwykłych ludzi. Przejrzystość jest tu jedyną odtrutką na lęk przed algorytmicznym Lewiatanem, który mógłby po cichu przemodelować nasze życie społeczne.

Siedemnastym filarem jest odporność na koncentrację, wymuszająca na nas refleksję nad tym, w czyich rękach spoczywa faktyczna władza nad technologią. AI rozwija się w świecie, w którym dostęp do modeli, chmury i talentów skupia się w rękach zaledwie kilku globalnych podmiotów, co rodzi ryzyko systemowe. Dojrzała strategia musi badać zjawisko vendor lock-in, dbać o interoperacyjność systemów oraz zapewniać realną możliwość migracji danych i procesów między różnymi dostawcami. Uzależnienie od jednej infrastruktury lub jednego, zamkniętego modelu sprawia, że organizacja staje się zakładnikiem cudzego regulaminu i cudzej polityki biznesowej. Warto pamiętać, że regulamin platformy nie jest konstytucją, choć czasem próbuje zachowywać się jak cesarz w kapciach, narzucając własne zasady bez możliwości negocjacji. Suwerenność cyfrowa wymaga więc dywersyfikacji i niezależnego audytu, aby fundamenty naszej wiedzy nie stały się własnością prywatnej korporacji.

Osiemnasty filar to współpraca interdyscyplinarna, która uznaje, że sztuczna inteligencja przestała być domeną wyłącznie informatyków i inżynierów danych. Specjaliści od kodu są oczywiście niezbędni, ale ich perspektywa jest niewystarczająca, by w pełni zrozumieć społeczne i prawne konsekwencje wdrażanych systemów. Potrzebujemy prawników, socjologów, psychologów, a nawet filozofów, którzy pomogą dostrzec ryzyka niewidoczne z poziomu czystej optymalizacji matematycznej. Model może być przecież technicznie poprawny i wysoce wydajny, a jednocześnie prawnie wadliwy, społecznie szkodliwy i etycznie nieakceptowalny. Tylko prawdziwa interdyscyplinarność pozwala zobaczyć tę nieprzyjemną prawdę w całości, zanim system zostanie uruchomiony na masową skalę. Nie chodzi tu o wygłaszanie mglistych uwag, lecz o twardą, wielostronną analizę, która chroni nas przed technokratyczną krótkowzrocznością.

Dziewiętnasty filar to odwaga odmowy, która w epoce wszechobecnej automatyzacji staje się jedną z najważniejszych kompetencji strategicznych każdej nowoczesnej instytucji. Organizacja musi umieć powiedzieć „nie” wdrożeniu systemów, które naruszają prywatność, pogłębiają nierówności lub są po prostu zbyt ryzykowne w danym kontekście. Jest to zadanie niezwykle trudne, zwłaszcza gdy konkurencja wydaje się mniej skrupulatna i szybciej eksploatuje nowe możliwości technologiczne. Jednak każda dojrzała cywilizacja poznaje samą siebie po tym, z czego potrafi zrezygnować mimo technicznej możliwości działania. Pytanie „czy możemy to zbudować?” jest domeną inżyniera, ale pytanie „czy powinniśmy to wdrożyć?” jest testem dojrzałości dla całej instytucji. W AI niestety wciąż zbyt często testuje się skrzydła na społeczeństwie, zapominając, że nie każdy upadek kończy się jedynie miękkim lądowaniem w arkuszu kalkulacyjnym.

Dwudziesty filar, domykający tę strukturę, to ciągłe uczenie się i zdolność do adaptacji w świecie, który nieustannie zmienia swoje parametry. Odpowiedzialna sztuczna inteligencja nie posiada końcowego stanu doskonałości, ponieważ modele, przepisy i normy społeczne będą ewoluować szybciej niż nasze procedury. System zarządzania musi być więc adaptacyjny, co oznacza zdolność do szybkiej korekty kursu na podstawie audytów, incydentów oraz skarg użytkowników. Każdy błąd powinien wpływać na architekturę systemu, a każde nowe badanie naukowe musi znajdować odzwierciedlenie w programach szkoleniowych dla pracowników. Dojrzała instytucja rozumie, że budowanie cyfrowej republiki to proces ciągły, wymagający czujności i gotowości do przyznania się do błędu. Tylko w ten sposób możemy uniknąć algorytmicznego feudalizmu, w którym dane i algorytmy organizują hierarchię społeczną bez żadnej realnej odpowiedzialności ze strony ich twórców.

Odpowiedzialność instytucji: między technologią a cywilizacją

Organizacja, która wdrożyła systemy sztucznej inteligencji i uznała to zadanie za definitywnie zakończone, przypomina chirurga, który po skomplikowanej operacji opuszcza salę, rzucając pacjentowi na odchodnym: „rana z pewnością sama zrozumie naszą strategię rekonwalescencji". Taki model dojrzałej instytucji AI, czyli organizacji traktującej systemy decyzyjne jako byty wymagające ciągłego nadzoru, a nie tylko jednorazowe narzędzia produktywności, wciąż wydaje się wielu decydentom egzotyczną utopią. Jest to droga trudna, kosztowna i pozbawiona medialnego blichtru, niemniej stanowi jedyną realną alternatywę dla automatyzacji wyzutej z odpowiedzialności. Bez tego fundamentu grozi nam bowiem regres do nowej, cyfrowej odmiany feudalizmu poznawczego, gdzie technologia, zamiast wyzwalać, staje się narzędziem nowej formy poddaństwa. W klasycznym feudalizmie to ziemia, przemoc i osobista lojalność organizowały strukturę władzy, wyznaczając granice między panem a chłopem. Dziś tę rolę przejmują dane, modele, infrastruktura oraz pogłębiająca się asymetria zrozumienia mechanizmów rządzących naszą rzeczywistością.

Feudalizm algorytmiczny, czyli model sprawowania władzy, w którym dane i algorytmy organizują hierarchię społeczną bez realnej odpowiedzialności twórców, staje się naszą codziennością. W tym nowym porządku jedni posiadają systemy klasyfikujące świat, podczas gdy drudzy są przez te systemy bezlitośnie klasyfikowani, często bez prawa do apelacji. Jedni rozumieją ukryte mechanizmy predykcji, drudzy zaś otrzymują jedynie suchy wynik, który determinuje ich szanse życiowe. Jedni optymalizują procesy dla zysku lub efektywności, drudzy są optymalizowani jako

zasoby w arkuszu kalkulacyjnym. To nie jest odległa wizja z literatury science-fiction, lecz bardzo prawdopodobna ścieżka rozwoju, jeśli etyka zostanie zredukowana do efektownej prezentacji w PowerPoincie, a audyt do formalnej pieczątki. Warto przy tym precyzyjnie odróżnić odpowiedzialną sztuczną inteligencję od tej, którą roboczo moglibyśmy nazwać AI „grzeczną".

Grzeczny model to taki, który nie używa wulgaryzmów, uprzejmie odmawia generowania oczywistych nadużyć i przemawia do nas kojącym, syntetycznym głosem. Odpowiedzialny system jest jednak czymś znacznie większym i bardziej skomplikowanym: posiada wbudowane ograniczenia władzy, przejrzystą ścieżkę audytu oraz rygorystyczne procedury korekty błędów. Wymaga on pełnej dokumentacji, stałego nadzoru ludzkiego, świadomości kontekstu kulturowego oraz mechanizmów, które pozwalają mu na przyznanie się do własnej niepewności. Grzeczność interfejsu bywa wręcz niebezpieczna, jeśli służy jedynie jako estetyczna maska dla głębokiej nieodpowiedzialności całego systemu. Miły chatbot, który z nienaganną kulturą odmawia wyjaśnienia przyczyn odrzucenia wniosku kredytowego, nadal pozostaje twarzą systemowej niesprawiedliwości, tyle że bardziej uprzejmą. Cywilizacja nie potrzebuje przede wszystkim maszyn o dobrych manierach, lecz rozliczalnych instytucji, które potrafią wziąć odpowiedzialność za ich działanie.

Z tej perspektywy książka redagowana przez Terezę Raquel Merlo i Jaya Liebowitza broni stanowiska, które jest zarazem zdumiewająco umiarkowane i głęboko radykalne. Umiarkowanie objawia się w tym, że autorzy nie odrzucają technologii a priori, nie popadają w jałową technofobię i nie wzywają do romantycznej ucieczki z nowoczesności. Radykalizm polega natomiast na kategorycznej odmowie traktowania etyki jako opcjonalnego dodatku do gotowego produktu. Etyka, audyt, infrastruktura oraz zarządzanie wiedzą są tutaj traktowane jako rdzeń systemu, a nie jego powierzchowna dekoracja czy hasło w strategii CSR. To przesłanie należy promować z całą stanowczością właśnie dlatego, że jest ono skrajnie niewygodne dla wszystkich stron ideologicznego sporu. Technoutopista uzna je za niepotrzebny hamulec postępu, technocynik za przejaw naiwności, a biurokrata spróbuje zamienić je w kolejny martwy formularz do odhaczenia.

Odpowiedzialny praktyk dostrzeże jednak w tym podejściu mapę przetrwania w coraz bardziej nieprzewidywalnym środowisku technologicznym. Przyszłość sztucznej inteligencji dla dobra wspólnego zależy od tego, czy potrafimy zbudować instytucje zdolne do uczenia się szybciej, niż rośnie ryzyko generowane przez ich własne narzędzia. To zdanie może brzmieć jak marketingowy slogan, ale w rzeczywistości jest zimną, niemal kliniczną diagnozą naszych czasów. Modele będą stawać się coraz tańsze, szybsze i bardziej autonomiczne, co doprowadzi do ich głębszej integracji z kluczowymi procesami decyzyjnymi państw i korporacji. Fałszywe treści staną się bardziej realistyczne, ataki socjotechniczne bardziej spersonalizowane, a pokusa całkowitej automatyzacji

będzie niemal niemożliwa do odparcia. W takim świecie nierówności kompetencyjne mogą gwałtownie rosnąć, a zaufanie publiczne do jakichkolwiek instytucji może ostatecznie lec w gruzach.

W świecie zdominowanym przez algorytmy nie wystarczy już samo posiadanie szczytnych zasad; kluczowa staje się zdolność do ich operacjonalizacji w codziennej praktyce. Moralność pozbawiona procedury niechybnie przegra w starciu z bezduszną automatyzacją, podczas gdy procedura wyzuta z moralności stanie się biurokratycznym pancerzem krępującym człowieka. Najważniejszym słowem w naszym słowniku nie powinna być zatem „inteligencja", lecz właśnie „odpowiedzialność", która definiuje granice dopuszczalnego działania. Inteligencja systemu mówi nam jedynie, co maszyna potrafi zrobić, natomiast odpowiedzialność instytucji określa, czego robić jej pod żadnym pozorem nie wolno. AI może być genialnym narzędziem służącym dobru wspólnemu w medycynie, finansach czy administracji, ale tylko pod warunkiem ścisłego powiązania z prawem i kulturą krytyczną. Inaczej stanie się jedynie narzędziem przyspieszenia wszystkiego, włączając w to ludzką głupotę, przemoc proceduralną oraz cyniczną propagandę.

W tym kontekście powraca do nas Henry David Thoreau, już nie jako ozdobny cytat z klasyki literatury, lecz jako twarde kryterium cywilizacyjne. Filozof przypominał, że nie wystarczy być pracowitym, wszak mrówki również są niezwykle pracowite, i stawiał fundamentalne pytanie: nad czym tak naprawdę się trudzisz? To samo pytanie należy dziś zadać każdemu modelowi, każdemu zarządowi korporacji, każdemu regulatorowi i każdemu państwu wdrażającemu systemy AI. Nad czym pracuje wasza maszyna: nad postawieniem lepszej diagnozy medycznej czy nad znalezieniem tańszego sposobu na odmowę należnej pacjentowi opieki? Czy budujecie systemy dla sprawiedliwszego podziału kredytów, czy może dla bardziej eleganckiego i niewidocznego wykluczenia całych grup społecznych? Czy wasza technologia służy demokracji bardziej inkluzywnej, czy może precyzyjniejszej propagandzie, która skuteczniej manipuluje emocjami wyborców? Czy stawiacie na edukację krytyczną, czy na produkcję gotowych, bezmyślnych odpowiedzi?

Dochodzimy tutaj do sedna problemu: przyszłość odpowiedzialnej sztucznej inteligencji nie zostanie rozstrzygnięta przez samą technologię, lecz przez instytucje. Muszą one zdecydować, czy chcą być jedynie szybsze w swoich działaniach, czy także mądrzejsze w swoich dalekosiężnych wyborach. Szybkość pozbawiona mądrości jest bowiem formą cywilizacyjnego poślizgu, a efektywność bez sprawiedliwości to nic innego jak dobrze zarządzana krzywda. Przejrzystość bez realnej możliwości sprzeciwu jest jedynie wystawą, a nie fundamentem demokracji, podobnie jak audyt bez konsekwencji staje się pustym teatrem. Dojrzała instytucja AI musi zatem połączyć pięć porządków: techniczny, prawny, ekonomiczny, społeczny oraz antropologiczny, aby zachować

cywilizacyjną spójność. Porządek techniczny zapewnia bezpieczeństwo i wyjaśnialność, prawny wyznacza granice i obowiązki, zaś ekonomiczny pozwala zrozumieć bodźce sterujące rozwojem systemów.

Porządek społeczny ukazuje nam wpływ technologii na wspólnotę, natomiast antropologiczny przypomina, że człowiek nie jest jedynie użytkownikiem, rekordem w bazie czy profilem marketingowym. Jest on istotą zdolną do nadawania sensu, do błędu, do uczenia się, a przede wszystkim do posiadania niezbywalnej godności. Sztuczna inteligencja, która o tym zapomina, może być technicznie genialna, ale nigdy nie będzie mogła zostać uznana za cywilizowaną. Nasza nowa maksyma powinna brzmieć: nie pytajmy, czy AI zastąpi człowieka, lecz czy instytucje używające AI nadal będą godne miana ludzkich. Jest to zadanie trudniejsze, mniej medialne i znacznie ważniejsze, ponieważ jeśli zbudujemy systemy efektywne kosztem słabszych, to problemem nie będzie maszyna. Problemem okaże się nasza stara skłonność do oddawania władzy temu, co obiecuje nam wygodę za cenę rezygnacji z sumienia.

Sztuczna inteligencja dla dobra wspólnego jest możliwa, ale nie powstanie ona samoistnie z mocy obliczeniowej coraz większych modeli językowych. Wymaga ona audytu jako konstytucji systemu, infrastruktury jako pamięci odpowiedzialności oraz zarządzania wiedzą jako pomostu między doświadczeniem a decyzją. Potrzebujemy etyki traktowanej jako architektura, edukacji jako świadomości krytycznej oraz demokracji jako wspólnej troski o kształt naszej rzeczywistości. Dopiero wtedy AI przestaje być tylko maszyną do przewidywania następnego tokenu czy kliknięcia, a staje się narzędziem społeczeństwa, które nie wyrzekło się rozumu. Społeczeństwo, które nie wyrzeka się rozumu i potrafi narzucić technologii ramy moralne, ma jeszcze realną szansę, by nie zostać zoptymalizowanym do własnej klęski. Etyka bez infrastruktury pozostanie jedynie kazaniem do serwera, podczas gdy my potrzebujemy fundamentów, na których zbudujemy przyszłość godną człowieka. Człowiek w pętli bez władzy jest przecież tylko dekoracją sumienia, a my potrzebujemy realnego sprawstwa.

Chmura danych jako fundament suwerenności państwa

Polska chmura danych to dzisiaj znacznie więcej niż ambitny projekt informatyczny; to fundamentalny sprawdzian suwerenności państwa w epoce cyfrowej, w której granice wyznaczają nie zasieki, lecz protokoły przesyłu informacji. W aparacie pojęciowym Teresy Raquel Merlo pytanie o chmurę nie jest bowiem kwestią techniczną, lecz stricte ustrojową, dotykającą samych fundamentów współczesnej władzy. Chmura w tym ujęciu przestaje być fizycznym „miejscem”, a staje się dynamicznym modelem przetwarzania, zależności oraz wielopoziomowej kontroli nad

zasobami informacyjnymi. Polskie Standardy Cyberbezpieczeństwa Chmur Obliczeniowych (SCCO) ujmują tę intuicję z niemal matematyczną precyzją, wskazując, że nowoczesne modele przetwarzania są całkowicie niezależne od fizycznego miejsca przechowywania bitów. To nie lokalizacja serwerowni, lecz architektura zarządzania danymi decyduje o tym, kto w sytuacjach kryzysowych trzyma rękę na pulsie państwowej infrastruktury.

To jedno zdanie o prymacie modelu nad miejscem powinno zostać wyryte nad drzwiami każdego ministerstwa, samorządu, banku czy szpitala, który właśnie wdraża systemy cyfrowe. Skoro bowiem chmura nie jest punktem na mapie, lecz modelem relacji, to suwerenność nad danymi nie sprowadza się do naiwnego pytania, czy serwer stoi pod Warszawą, we Frankfurcie czy w Wirginii. Prawdziwa stawka dotyczy tego, kto sprawuje faktyczną władzę nad dostępem, kluczami szyfrującymi, personelem operacyjnym oraz prawem właściwym dla rozstrzygania sporów. Z perspektywy Merlo chmura państwowa to nie jest po prostu cyfrowy magazyn plików, lecz skomplikowana infrastruktura zaufania, na której opiera się stabilność instytucji. Dane państwa, biznesu i obywateli są dzisiaj surowcem administracji, paliwem dla sztucznej inteligencji, a w razie konfliktu – głównym przedmiotem cyberwojny i materiałem do głębokiego profilowania społecznego.

Kto kontroluje te dane, ten nie tylko „zarządza informacją" w tradycyjnym, biurowym sensie tego słowa, lecz współdecyduje o tym, co państwo widzi i jak szybko potrafi zareagować na zagrożenie. Kontrola nad chmurą determinuje, kogo systemy klasyfikują jako ryzyko, komu algorytm przyzna świadczenie, a kogo podda drobiazgowej kontroli skarbowej. W tym kontekście polska debata o chmurze jawi się jako spóźniona, lecz absolutnie konieczna operacja na otwartym, cyfrowym kręgosłupie państwa. Polskie decyzje strategiczne w tej materii układają się w kilka warstw, z których pierwszą jest inicjatywa Wspólna Infrastruktura Informatyczna Państwa (WIIP). Jest to próba zbudowania spójnych zasad korzystania z Rządowej Chmury Obliczeniowej oraz Publicznej Chmury Obliczeniowej, co ma wreszcie ujednolicić standardy bezpieczeństwa dla całej administracji.

Standardy SCCO wskazują, że strategiczne kierunki WIIP obejmują nie tylko podniesienie bezpieczeństwa usług elektronicznych, ale także radykalne obniżenie kosztów stałych i zwiększenie efektywności wydatkowania środków na IT. Chodzi o skrócenie czasu realizacji projektów informatycznych oraz ograniczenie patologii wielokrotnego gromadzenia tych samych danych przez różne urzędy. Choć brzmi to jak typowy język modernizacji technokratycznej, pod spodem kryje się proces znacznie poważniejszy: państwo próbuje przestać być zbiorem rozproszonych, analogowych silosów i stać się spójnym organizmem cyfrowym. Druga warstwa to polityka suwerenności danych, czyli pełna prawna i operacyjna kontrola nad informacjami, niezależna od jurysdykcji państw

trzecich. Ministerstwo Cyfryzacji w marcu 2025 roku zaczęło mówić o tym wprost, definiując rozwój chmury jako fundament bezpieczeństwa narodowego.

Ten zwrot w języku oficjalnym jest niezwykle istotny, ponieważ przesuwa punkt ciężkości z czystej efektywności ekonomicznej na twardą, polityczną kontrolę. Jeszcze kilka lat temu opowieść o chmurze była przede wszystkim radosną narracją o skalowalności, oszczędnościach i wygodzie pracy w modelu DevOps. Dziś jest to opowieść o suwerenności, ekstraterytorialności prawa, cyberodporności oraz realnym ryzyku strategicznego szantażu ze strony globalnych dostawców. Trzecia warstwa to praktyka partnerstw publiczno-prywatnych, która pokazuje, że Polska nie buduje swojej cyfrowej twierdzy w całkowitej izolacji. Operator Chmury Krajowej (OChK) działa jako wyspecjalizowany integrator, oferując zarówno własną platformę, jak i rozwiązania globalnych gigantów, takich jak Google Cloud czy Microsoft Azure.

W lutym 2025 roku PFR, OChK i Google Cloud Poland podpisały memorandum dotyczące strategicznego partnerstwa w zakresie wdrażania rozwiązań AI w polskich instytucjach. Jest to model hybrydowy, w którym Polska próbuje połączyć krajowe kompetencje z potęgą technologiczną globalnych hyperscalerów. Takie podejście jest pragmatyczne, ale jednocześnie nasycone napięciami, których nie da się łatwo rozładować marketingowym optymizmem. Tutaj powraca zasadniczy problem metodologiczny Merlo: technologia zaufana to nie ta, która brzmi dumnie i narodowo, lecz ta, której łańcuch odpowiedzialności jest w pełni audytowalny. „Chmura krajowa" nie staje się automatycznie suwerenna tylko dzięki swojej nazwie, tak jak „etyczna sztuczna inteligencja" nie staje się etyczna tylko dlatego, że ktoś umieścił takie hasło na slajdzie w programie PowerPoint.

W procesie humanizacji technologii musimy pytać o rzeczy fundamentalne: kto jest faktycznym właścicielem infrastruktury i kto administruje kluczami szyfrującymi? Jakie obce prawo może wymusić dostęp do naszych danych i gdzie fizycznie znajduje się personel operacyjny czuwający nad systemem w nocy? Należy badać procedury incydentowe, dostęp do metadanych oraz to, czy państwo posiada realną ścieżkę wyjścia, czy może wpadło w pułapkę vendor lock-in, czyli całkowitego uzależnienia od technologii jednego dostawcy. Kluczowe jest również to, czy w razie awarii, ataku lub nagłego sporu prawnego administracja potrafi samodzielnie odtworzyć działanie krytycznych systemów. Bez tej wiedzy suwerenność cyfrowa pozostaje jedynie publicystycznym ozdobnikiem, a nie realnym atrybutem władzy państwowej.

Polska, podobnie jak większość państw średniej wielkości, stoi dziś przed dylematem zależności inteligentnej. Pełna autarkia chmurowa, czyli próba zbudowania wszystkiego od zera własnymi siłami, jest dziś mało realistyczna, astronomicznie kosztowna i technologicznie ryzykowna. Z drugiej strony, bezrefleksyjna zależność od amerykańskich Big Techów jest strategicznie

lekkomyślna i wystawia nas na łaskę obcych interesów korporacyjnych. Dlatego rozsądna ścieżka nie polega ani na cyfrowym nacjonalizmie z tektury, ani na chmurowym liberalizmie bez żadnych granic. Rozwiązaniem jest model warstwowy: dane i systemy o najwyższym stopniu wrażliwości muszą pozostać pod ścisłą kontrolą jurysdykcyjną i techniczną państwa, podczas gdy usługi powszechne mogą korzystać z chmur publicznych.

Problem polega jednak na tym, że klasyfikacja danych i systemów jest w praktyce znacznie trudniejsza, niż mogłoby się wydawać na pierwszy rzut oka. Dane samorządowe o gospodarce odpadami wydają się na pozór niewinne, dopóki nie zostaną połączone z informacjami o lokalizacji, płatnościach i zużyciu mediów, tworząc precyzyjny obraz życia obywateli. Dane edukacyjne wyglądają na czystą administrację, dopóki nie zaczną służyć do budowania algorytmicznych profili dzieci, determinujących ich przyszłe szanse życiowe. Nawet dane o nieobecnościach w pracy czy zwolnieniach lekarskich, połączone z historią zatrudnienia, mogą prowadzić do niezwykle głębokiego profilowania jednostki. Metodologia Merlo podpowiada nam tutaj: nie klasyfikuj danych wyłącznie po nazwie rejestru, lecz po możliwej szkodzie, jaką może wyrządzić ich wyciek lub niewłaściwe użycie algorytmiczne. Chmura nie jest niebem; chmura jest cudzym komputerem, cudzym prawem i cudzym modelem biznesowym, chyba że sami zbudujemy w niej instytucję odpowiedzialności.

Między suwerennością a zależnością: lekcje dla cyfrowej Polski

Rezydencja danych to w gruncie rzeczy tylko adres na mapie, fizyczne „tu i teraz” bitów zapisanych na konkretnym dysku w określonej lokalizacji. Suwerenność danych, czyli pełna prawna i operacyjna kontrola nad informacją niezależnie od miejsca jej składowania, to już zupełnie inna kategoria wagowa, wymagająca nie tylko kabli, ale i solidnych paragrafów. Lokalizacja danych to z kolei administracyjny wymóg, by określone zasoby nie opuszczały terytorium państwa, co w dobie globalnych sieci bywa postulatem tyleż słusznym, co technicznie karkołomnym. Dane mogą przecież fizycznie leżeć w serwerowni pod Warszawą, a mimo to pozostawać pod przemożnym wpływem prawa państwa trzeciego, jeśli ich dostawca jest częścią grupy kapitałowej podlegającej ekstraterytorialnym przepisom. Francuska debata o SecNumCloud i amerykańskim CLOUD Act postawiła ten problem na ostrzu noża, uświadamiając Europie, że zgodność z RODO nie jest automatycznie równoznaczna z pełną suwerennością. Chmura nie jest niebem; chmura jest cudzym komputerem, cudzym prawem, cudzym personelem i cudzym modelem biznesowym, o czym często zapominamy w technokratycznym uniesieniu.

Francuskie rozwiązania typu „cloud de confiance” to ambitna próba zabezpieczenia nie tylko fizycznej lokalizacji serwerów, lecz przede wszystkim odporności na obce żądania dostępu do informacji. Paryż nie zadowolił się ogólnikowym językiem cyfryzacji, lecz stworzył spójną doktrynę chmury zaufanej, rozwijając rygorystyczne wymagania certyfikacji SecNumCloud. Promuje ona modele hybrydowe, takie jak S3NS, owoc współpracy Thalesa z Google Cloud, czy Bleu, łączący siły Orange, Capgemini i Microsoftu. Nie mamy tu do czynienia z czystym modelem narodowym, lecz z próbą „udomowienia” amerykańskiej technologii w strukturach prawnych kontrolowanych przez podmioty francuskie. Decyzja o przeniesieniu Health Data Hub z Microsoft Azure do rodzimego Scaleway pokazuje, że obawa przed ekstraterytorialnym dostępem do danych zdrowotnych milionów obywateli stała się twardą decyzją państwową, a nie tylko publicystyczną fobią. Francuska lekcja brzmi donośnie: suwerenność chmurowa nie jest wyrazem antyamerykanizmu, lecz dojrzałym zarządzaniem ryzykiem strategicznej zależności.

Państwo może i powinno współpracować z Big Techami, ale musi też umieć wyraźnie wskazać punkt, w którym kończy się wygodny outsourcing, a zaczyna nienaruszalny rdzeń zaufania publicznego. Polska powinna tę lekcję czytać bez kompleksów, ale i bez imitacyjnego entuzjazmu, który każe nam kopiować rozwiązania bez uwzględnienia lokalnego kontekstu. Francja dysponuje większym rynkiem, silniejszymi czempionami technologicznymi i specyficzną, scentralizowaną kulturą strategiczną, której nie da się przenieść nad Wisłę jeden do jednego. Możemy jednak przyjąć francuską metodę: klasyfikować dane według ich wrażliwości, wymagać pełnej przejrzystości łańcucha dostaw i certyfikować środowiska chmurowe z najwyższą starannością. Najważniejsze to przestać mylić „serwer w Europie” z realną „kontrolą europejską”, bo to dwa zupełnie inne porządki bezpieczeństwa. Etyka bez infrastruktury jest przecież tylko kazaniem do serwera, który i tak nas nie słucha, jeśli nie my trzymamy klucze do jego obudowy.

Chiny idą zupełnie inną drogą, budując swoją potęgę nie przez hybrydowe partnerstwa, lecz przez bezwzględną logikę cyber-suwerenności i dominację własnych gigantów, takich jak Alibaba Cloud czy Huawei. Chiński model jest infrastrukturalnie skuteczny, ale politycznie opiera się na absolutnym prymacie państwa nad wolnością jednostki i społeczeństwem obywatelskim. Tamtejsze prawo nakłada na firmy szerokie obowiązki współpracy z organami bezpieczeństwa, co czyni z danych narzędzie zarówno gospodarczej ekspansji, jak i politycznej dyscypliny. Dla Polski nie jest to wzór do naśladowania, lecz memento: suwerenność danych bez silnych praw obywatelskich może łatwo przeobrazić się w suwerenność władzy nad obywatelem. W takim układzie dane nie służą już wspólnocie, lecz stają się paliwem dla cyfrowego Lewiatana, który nie znosi sprzeciwu ani prywatności.

Rosja reprezentuje jeszcze ostrzejszą, niemal brutalną wersję lokalizacyjnego etatyzmu, wymagając bezwzględnego przechowywania danych osobowych swoich obywateli na terytorium kraju. Zmiany w prawie zaplanowane na 2025 rok jeszcze bardziej zaostrzają te rygory, wprowadzając drakońskie sankcje za niepodporządkowanie się woli Kremla. Formalnie jest to polityka suwerenności, ale w systemie autorytarnym służy ona przede wszystkim ułatwieniu inwigilacji i sprawnej represji niepokornych jednostek. Tu znów widać metodologiczną przewagę podejścia, w którym nie wolno oceniać suwerenności wyłącznie po tym, czy dane fizycznie znajdują się „u nas". Kluczowe pytanie brzmi: u kogo właściwie jest to „u nas", pod czyją kontrolą operacyjną i z jakimi gwarancjami procesowymi dla zwykłego człowieka? Bez niezależnych sądów i audytów, lokalizacja danych staje się jedynie logistycznym ułatwieniem dla policji politycznej.

Izrael pokazuje nam trzeci wariant: państwo wysokich technologii, które traktuje chmurę jako integralny element architektury bezpieczeństwa narodowego i negocjuje z Big Techami z pozycji siły. Projekt Nimbus, czyli gigantyczny kontrakt rządu z Google i Amazonem, był przedstawiany jako cywilizacyjny skok, ale szybko obnażył brutalną prawdę o cyfrowych sojuszach. Doniesienia z 2025 roku sugerują istnienie tajnych mechanizmów informowania państwa o żądaniach dostępu do danych przez zagraniczne organy, co czyni z dostawców niemal agentów wywiadu. Izraelski przypadek ma także mroczny wymiar etyczny, o czym świadczy odcięcie przez Microsoft dostępu jednostki 8200 do usług Azure po oskarżeniach o masową inwigilację Palestyńczyków. To podręcznikowy przykład na to, że technologia nigdy nie jest neutralna, jeśli służy systemowi przemocy, a dostawca mocy obliczeniowej staje się współodpowiedzialny za skutki jej użycia.

Polska musi zatem rozpoznać własną drogę między tymi modelami, unikając zarówno naiwnej autarkii, jak i bezwarunkowej zależności od globalnych potęg. Powinniśmy budować model demokratycznej suwerenności danych, oparty na audytowalnej kontroli nad zasobami krytycznymi i realnych krajowych kompetencjach technicznych. Mamy już pewne ramy formalne, od RODO po ustawę o krajowym systemie cyberbezpieczeństwa, ale sama obecność przepisów nie oznacza jeszcze dojrzałego zarządzania. Administracja publiczna idzie w stronę cyfryzacji usług i szerokiego wykorzystania sztucznej inteligencji, co widać w ambitnych planach rozwoju AI do 2030 roku. Jednak bez jawności i kontroli społecznej, te nowoczesne narzędzia mogą stać się fundamentem pod feudalizm algorytmiczny, czyli model władzy, w którym algorytmy organizują hierarchię społeczną bez realnej odpowiedzialności ich twórców.

W polskim systemie wciąż brakuje pełnej, publicznej mapy algorytmów używanych przez państwo do podejmowania decyzji, typowania kontroli czy priorytetyzacji spraw obywateli. Wiedza o tym, jak działają te cyfrowe wyrocznie, jest fragmentaryczna i często pochodzi z raportów organizacji społecznych, a nie z przejrzystych rejestrów publicznych. Raporty typu AlgoPolska od lat ostrzegają,

że zautomatyzowane podejmowanie decyzji w sferze publicznej, choć efektywne, niesie ze sobą ryzyko błędów, nierówności i braku przejrzystości. Przykładem jest ZUS i jego narzędzia analityczne używane do typowania zwolnień lekarskich do kontroli, gdzie brakuje jasnych kryteriów i możliwości realnego zakwestionowania wyników pracy maszyny. Człowiek w pętli bez władzy jest tylko dekoracją sumienia, a my potrzebujemy systemów, które wzmacniają obywatela, a nie tylko ułatwiają pracę urzędowi.

Problem nie polega na tym, że państwo używa zaawansowanej analityki do wykrywania nadużyć, bo to działanie w interesie publicznym i przejaw nowoczesności. Prawdziwym wyzwaniem jest ukryta translacja polityki, czyli proces zamiany ogólnych celów społecznych na konkretne modele predykcyjne bez jawnego audytu i wyjaśnienia ich mechanizmów. Jeśli nie wiemy, dlaczego algorytm uznał nas za podejrzanych, tracimy podmiotowość w relacji z państwem, które staje się nieprzeniknioną, cyfrową ścianą. Polska potrzebuje zatem wizji cyfrowej republiki, w której architektura chmurowa i algorytmiczna służy realizacji gwarancji konstytucyjnych, a nie tylko optymalizacji procesów biurokratycznych. Suwerenność danych musi oznaczać przede wszystkim suwerenność obywatela nad własną informacją, chronioną przez silne państwo przed zakusami zarówno globalnych korporacji, jak i pokusami własnej administracji. Tylko wtedy cyfryzacja nie będzie kolejnym etapem budowy „cesarstwa w kapciach", lecz realnym wzmocnieniem demokratycznego państwa prawa.

Państwo prawa w cieniu algorytmu: wyzwania transparentności

Państwo posiada prawo do ochrony funduszy publicznych, niemniej kluczowe pytanie brzmi, czy obywatel wie, że został wytypowany przez system. Problem dotyczy nie tylko rodzaju przetwarzanych danych, ale tego, czy algorytm nie faworyzuje określonych profili bez realnego nadzoru człowieka. W metodologii Merlo zjawisko to określa się mianem ukrytej translacji polityki publicznej, czyli procesu zamiany celów politycznych na modele predykcyjne bez jawnego audytu. Kiedy polityk postuluje walkę z nadużyciami, a urzędnik optymalizację kontroli, analityk buduje model ryzyka, który obywatel odczuwa jako nagłe wezwanie lub odmowę. Między szczytnym hasłem a doświadczeniem jednostki zachodzi algorytmiczna translacja wartości, która bezwzględnie powinna podlegać weryfikacji. Inaczej państwo powie „sprawiedliwość", system zakoduje „podejrzliwość", a obywatel otrzyma jedynie chłodną kontrolę.

W polskim kontekście należy zachować ostrożność, bowiem brakuje twardych dowodów na to, że strategiczne decyzje rządu są podejmowane przez autonomiczne algorytmy. Wiemy jednak, iż

administracja rozwija narzędzia analityczne, a oficjalna polityka AI przewiduje wsparcie procesów legislacyjnych przez systemy uczenia maszynowego. Partie polityczne, wzorem globalnych trendów, korzystają z analityki danych do monitorowania opinii oraz precyzyjnego targetowania komunikacji w mediach społecznościowych. Warto zatem rozróżnić algorytmy jako narzędzia wykonawcze, instrumenty analizy politycznej oraz potencjalne mechanizmy faktycznego współdecydowania. Ten ostatni poziom, będący formą cyfrowej republiki, czyli wizji państwa realizującego gwarancje konstytucyjne poprzez technologię, wymagałby najwyższego stopnia transparentności.

Deficytem Polski nie jest samo wykorzystanie technologii, lecz brak pełnej instytucjonalizacji jawności algorytmicznej w strukturach państwowych. Państwo powinno prowadzić publiczny rejestr systemów używanych w administracji, z wyłączeniem jedynie obszarów wymagających szczególnego bezpieczeństwa. Przy czym rejestr musiałby obejmować cel systemu, kategorie danych, podstawę prawną, rolę człowieka oraz szczegółowe informacje o przeprowadzonych audytach. Bez jasnych procedur odwoławczych obywatel nie wie, kiedy rozmawia z państwem, a kiedy jedynie z jego cyfrowym modelem. Państwo prawa nie powinno nosić maski interfejsu, ukrywając mechanizmy władzy za nieprzeniknioną ścianą kodu i skomplikowanych algorytmów.

W przypadku samorządów problem jest rozproszony, gdyż przetwarzają one dane o edukacji, pomocy społecznej czy planowaniu przestrzennym przy ograniczonych zasobach. Lokalne wspólnoty dysponują zazwyczaj mniejszymi budżetami i słabszymi zespołami cyberbezpieczeństwa, co czyni je podatnymi na dyktat zewnętrznych dostawców. Migracja do chmury może poprawić ciągłość działania, jednak bez odgórnych standardów samorządy ryzykują stanie się archipelagiem zależności od dostawców. Państwo musi zapewnić wsparcie w zakresie wspólnych zakupów i kontroli kontraktów, aby lokalna administracja zachowała suwerenność danych, czyli pełną prawną i operacyjną kontrolę nad informacjami niezależnie od miejsca ich przechowywania.

Biznes znajduje się w zróżnicowanym położeniu, gdzie wielcy gracze posiadają rozwinięte działy compliance, a mniejsze firmy często błądzą w chmurze. Statystyki pokazują, że w 2024 roku jedynie 5,9 procent polskich przedsiębiorstw korzystało z AI, przy średniej unijnej wynoszącej 13,5 procent. Polska musi nie tylko chronić dane, ale przede wszystkim podnieść kompetencje biznesu, aby technologia nie była jedynie importowaną przewagą. Małe firmy często migrują do rozwiązań chmurowych z powodów kosztowych, nie zawsze rozumiejąc konsekwencje prawne i jurysdykcyjne. Chmura nie jest niebem; jest cudzym komputerem, o czym przedsiębiorcy zbyt często zapominają.

Wrażliwe dane obywateli stanowią najważniejszy test moralny, ponieważ w tej relacji jednostka nie posiada realnej swobody negocjacyjnej. Wszak nikt nie może odmówić przekazania informacji do rejestru PESEL czy urzędu skarbowego tylko dlatego, że nie podoba mu się architektura chmury. Relacja obywatel–państwo jest relacją przymusowego zaufania, co nakłada na administrację

obowiązek zachowania wyższych standardów niż w zwykłym biznesie. Państwo nie może mówić „zaakceptuj regulamin albo nie korzystaj”, lecz musi gwarantować pełną audytowalność swoich systemów. Właśnie dlatego standard ochrony musi być tu bezkompromisowy, a każda decyzja algorytmiczna powinna być możliwa do wyjaśnienia i zakwestionowania.

Aparat Merlo daje silny argument, że dane wrażliwe wymagają etycznego zarządzania całym cyklem ich życia, a nie tylko technicznego zabezpieczenia. Musimy wiedzieć, po co dane są zbierane, jak długo przechowywane i czy służą one trenowaniu modeli bez naszej wiedzy. Niezbędne jest wprowadzenie mechanizmu decision provenance, czyli praktyki dokumentowania pochodzenia i logiki każdej decyzji, oraz zapewnienie realnej obecności człowieka w pętli decyzyjnej. Człowiek w pętli bez władzy jest tylko dekoracją sumienia, dlatego systemy muszą podlegać regularnym testom typu red teaming. Tylko poprzez pełną transparentność państwo może budować autentyczne zaufanie w erze algorytmicznej transformacji.

Cyfrowa republika: suwerenność danych w epoce chmury

To, co na pierwszy rzut oka wygląda na nudną, technokratyczną dłubaninę w serwerowniach, stanowi w istocie cyfrowy odpowiednik twardych gwarancji konstytucyjnych. Polska stoi dziś przed wyborem o charakterze strategicznym, który zdefiniuje naszą podmiotowość na dekady. Możemy potraktować chmurę państwową jedynie jako kolejny program modernizacji infrastruktury IT albo uczynić z niej fundament architektury cyfrowej republiki, czyli wizji państwa realizującego gwarancje konstytucyjne poprzez technologię. W pierwszym, dość leniwym wariancie, będziemy po prostu migrować systemy, kupować gotowe usługi i cieszyć się mityczną skalowalnością, podpisując kolejne memoranda z uśmiechem na twarzy. Prawdziwe wyzwanie leży jednak w drugim scenariuszu, w którym budujemy suwerenne środowiska dla danych krytycznych oraz jawny rejestr algorytmów.

Budowa cyfrowej republiki wymaga stworzenia precyzyjnej klasyfikacji danych według stopnia możliwej szkody społecznej oraz wdrożenia realnych procedur audytu technicznego. Kluczowe jest stworzenie mechanizmów odwoławczych dla obywateli, aby nikt nie stał się ofiarą feudalizmu algorytmicznego, czyli modelu władzy, w którym algorytmy organizują hierarchię społeczną bez realnej odpowiedzialności twórców. Potrzebujemy krajowych zdolności do przechowywania i przetwarzania informacji, które nie będą opierać się na całkowitej, niemal kolonialnej zależności od zewnętrznych korporacji. Wszak suwerenność danych, oznaczająca pełną prawną i operacyjną kontrolę nad zasobami niezależnie od jurysdykcji państw trzecich, nie jest jedynie technologicznym

luksusem, lecz wymogiem bezpieczeństwa. Niemniej, aby to osiągnąć, państwo musi przestać postrzegać cyfryzację wyłącznie przez pryzmat faktur i licencji.

Największa wątpliwość nie dotyczy bowiem samej technologii, lecz kondycji państwowej wyobraźni, która zbyt często grzęźnie w biurokratycznych procedurach. Czy Polska potrafi myśleć o danych jako o kluczowym zasobie ustrojowym, a nie tylko jako o uciążliwym problemie z zakresu zamówień publicznych? To pytanie o to, czy zbudujemy kadry zdolne negocjować z globalnymi hiperskalatorami jak równorzędny partner, a nie jak klient zachwycony sezonowym rabatem. Musimy wymagać pełnej audytowalności algorytmów nawet wtedy, gdy dany system działa wyjątkowo wygodnie dla administracji i ułatwia życie urzędnikom. Prawdziwym testem będzie stworzenie instytucji kontrolnej wyposażonej w realną wiedzę techniczną, a nie tylko w paragrafy i pieczątki.

Państwo musi umieć ochronić obywatela przed zjawiskiem automatyzacji podejrzenia, gdzie algorytm bezrefleksyjnie przypisuje winę na podstawie statystycznych korelacji. Istnieje bowiem ryzyko, że zaczniemy korzystać z AI w procesach legislacyjnych, zamieniając żywą politykę w syntetyczne uzasadnienia do decyzji, które zapadły już wcześniej. Taka ukryta translacja polityki, czyli proces zamiany celów państwa na konkretne modele predykcyjne bez jawnego audytu, stanowiłaby zaprzeczenie idei demokratycznego państwa prawa. Dlatego polska chmura musi być oceniana nie przez pryzmat patriotycznej retoryki, lecz według pięciu konkretnych kryteriów odpowiedzialnej AI. Obejmują one przede wszystkim realną kontrolę jurysdykcyjną nad danymi krytycznymi oraz pełną audytowalność decyzji wpływających na prawa obywatelskie.

Kryteria te wymagają także dojrzałego zarządzania wiedzą, które traktuje systemy AI jako instytucje decyzyjne wymagające nadzoru, a nie tylko jako narzędzia produktywności. Niezbędna jest demokratyczna przejrzystość oraz zdolność do stanowczej odmowy w relacjach z gigantami technologicznymi, co Francja uczyniła już w odniesieniu do wrażliwych danych zdrowotnych. Chiny i Rosja mówią o suwerenności w języku autorytarnej kontroli, podczas gdy Izrael definiuje ją przez pryzmat bezpieczeństwa militarnego. Polska powinna wypracować własny język, oparty na fundamentach państwa prawa, w którym współpraca z dostawcami jest środkiem, a nie celem. Chmura nie jest niebem, lecz cudzym komputerem i modelem biznesowym, chyba że sami zbudujemy w niej instytucję odpowiedzialności. Etyka bez infrastruktury jest przecież tylko kazaniem do serwera.

> Czy w pogoni za cyfrową wydajnością nie zamieniamy naszych instytucji w bezduszne serwery, które zamiast służyć ludziom, jedynie przetwarzają ich na dane? Prawdziwym sprawdzianem naszej cywilizacji nie jest szybkość wdrożenia algorytmów, lecz odwaga,

> by zatrzymać maszynę, gdy jej logika przestaje być ludzka. Pytanie brzmi: czy zbudujemy systemy, które wzmacniają naszą wolność, czy też staniemy się ostatecznie własną, zoptymalizowaną do klęski adaptacją?

Inspiracje

[1] **"Ethical AI and Data Science"** Jay Liebowitz, Tereza Raquel Merlo

2026 | Taylor and Francis | ISBN: 9781041110064

Jay Liebowitz to wybitny autorytet w dziedzinie sztucznej inteligencji (AI) i autor licznych publikacji. Jest profesorem innowacji biznesowych i dyrektorem Centrum AI-EDGE w Crummer Graduate School of Business na Rollins College. Wcześniej pełnił funkcje m.in. w NASA oraz na Johns Hopkins University. Jest również założycielem i redaktorem naczelnym czasopisma "Expert Systems With Applications".

*Dr. **Jay Liebowitz** is a distinguished academic and expert in artificial intelligence, knowledge management, and business innovation. Currently serving as a Professor of Business Innovation and Industry Transformation at Rollins College, he has held prominent positions at institutions including Johns Hopkins University, the University of Maryland, and the U.S. Army War College. Dr. Liebowitz is widely recognized for his pioneering work in expert systems and knowledge management, having served as the founding editor-in-chief of the journal 'Expert Systems With Applications'. With over 45 books and numerous scholarly articles to his credit, his research focuses on the intersection of technology, organizational intelligence, and ethical AI. He has also served as a Knowledge Management Officer at NASA and is a Fulbright Scholar, consistently contributing to the discourse on how digital transformation and intelligent systems impact society, governance, and business strategy.*

Rollins College

Tereza Raquel Merlo jest badaczką podoktorską na Federalnym Uniwersytecie Rio de Janeiro (UFRJ), afiliowaną przy Brazylijskim Instytucie Informacji w Nauce i Technologii (IBICT), oraz adiunktem na University of North Texas. Jej główne obszary działalności obejmują zarządzanie wiedzą, sztuczną inteligencję i analizę danych. Jest autorką książki „Understanding, Implementing, and Evaluating Knowledge Management in Business Settings” (2022) oraz redaktorką „Ethical AI and Data Science” (2026).

***Tereza Raquel Merlo** is a postdoctoral researcher at the Federal University of Rio de Janeiro (UFRJ), affiliated with the Brazilian Institute of Information in Science and Technology (IBICT), and an Adjunct Professor at the University of North Texas. Her main areas of activity include knowledge management, artificial intelligence, and data analytics. She is the author of "Understanding, Implementing, and Evaluating Knowledge Management in Business Settings" (2022) and editor of "Ethical AI and Data Science" (2026).*

Literatura uzupełniająca

Książki

Literatura pokrewna (Google Scholar):

[1] **"Ethical AI and Data Science: Building Trustworthy and Transparent Systems"**

Tereza Raquel Merlo, Jay Liebowitz

2026 | Routledge | ISBN: 978-1041109297

[2] **"Knowledge Retention: Strategies and Solutions"**

Jay Liebowitz

2009 | Taylor & Francis | ISBN: 978-1420064667

[3] **"Walden; or, Life in the Woods"**

Henry David Thoreau

1854 | Ticknor and Fields | ISBN: 978-0140390445

[4] **"Civil Disobedience"**

Henry David Thoreau

1849 | Elizabeth Peabody (Aesthetic Papers) | ISBN: 978-0486275635

[5] **"The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power"**

Shoshana Zuboff

2019 | PublicAffairs

[6] **"Digital Ethics"**

Christian Fuchs

2023 | Routledge

Artykuły naukowe

Autorzy przywoływani:

[1] **"Ethics of Artificial Intelligence: Issues and Initiatives"**

John C. Havens

2019 | IEEE Technology and Society Magazine

Google Scholar

[2] "Social and Philosophical Aspects of Digital Ethics: Challenges, Prospects, and Significance in the Technological Progress Era"

Marina V. Alekseeva, Anna A. Rokotyanskaya

2025 | Scilit

Google Scholar

Tekst opracowany z udziałem sztucznej inteligencji.
Redakcja i kontrola merytoryczna: Fundacja Dobre Państwo.
APA ONE Publishing System
Fundacja Dobre Państwo © 2026
Article ID: a479512a-8c2e-4388-8477-36c6983301e2