

NARZĘDZIA AI

Dla obywateli i administracji w epoce rewolucji agentowej



FORMUŁA NACZELNA:

W epoce rewolucji agentowej narzędzia AI budują państwo zdolne i obywatela podmiotowego wyłącznie jako system symetryczny i sprzężony: każdej zdolności AI administracji musi odpowiadać równoważna zdolność AI obywatela — pod mandatem człowieka i w jawnej pętli sprzężenia zwrotnego.

SYMETRIA NARZĘDZI AI: PAŃSTWO ZDOLNE — OBYWATEL PODMIOTOWY



Każdej zdolności AI administracji odpowiada równoważna zdolność AI obywatela — pod mandatem człowieka i w jawnej pętli sprzężenia zwrotnego.

Rysunek 1. Symetria narzędzi AI: agent państwa i agent obywatela sprzężone wspólną warstwą zaufania, danych i kontroli.

Formuła naczelna

Hasło programowe:

NIE CHATBOT W URZĘDZIE, LECZ SYMETRIA AGENTÓW.

NIE FASADA CYFROWA, LECZ SPRZĘŻONA PODMIOTOWOŚĆ.

Poprzednia wersja tej publikacji (sierpień 2026) opierała się na formule: „Narzędzia AI realizują algorytm służby obywatelom i administracji państwowej”. Formuła ta pozostaje prawdziwa jako cel, lecz przestała wystarczać jako reguła projektowa. W ciągu dwóch lat sztuczna inteligencja przeszła od narzędzia, które odpowiada na pytanie, do agenta, który samodzielnie planuje i wykonuje wieloetapowe zadania, korzysta z narzędzi, systemów i interfejsów oraz porozumiewa się z innymi agentami. Zmiana ta nie jest przyrostowa. Zmienia rozkład władzy między tym, kto dysponuje agentem, a tym, kto go nie ma.

Nowa formuła naczelna zawiera cztery rozstrzygnięcia. Każde z nich wyklucza konkretny błąd, który w cyfryzacji państwa powtarza się niezależnie od ekipy rządzącej i od kraju.

Rozstrzygnięcie	Treść	Wyklucza błąd
Rewolucja agentowa	AI działa, a nie tylko odpowiada; tempo zmiany mierzy się miesiącami	traktowanie AI jako kolejnego modułu informatycznego doklejonego do niezmienionego procesu
Symetria	każdej zdolności AI państwa odpowiada równoważna zdolność AI obywatela	asymetrię: państwo nadzoru po jednej stronie albo zalew wniosków generowanych przez agentów po drugiej
Sprzężenie	narzędzia obu stron łączy wspólna warstwa tożsamości, danych, formatów i zobowiązań	fasadę: rejestr, którego nikt nie czyta, i agenta, który „zgaduje” interfejs urzędu
Mandat człowieka	decyzje prawnie doniosłe i ustrojowe pozostają przy człowieku, zdolnym zrozumieć, zatrzymać i skorygować system	abdykację odpowiedzialności na dostawcę modelu albo na „wynik algorytmu”

REGUŁA WARUNKÓW ŁĄCZNYCH

Cztery człony formuły nie są wariantami do wyboru. Rewolucja agentowa bez symetrii prowadzi do państwa, które widzi obywatela lepiej niż obywatel widzi państwo. Symetria bez sprzężenia daje dwa równoległe światy narzędzi, które się nie rozumieją. Sprzężenie bez mandatu człowieka tworzy zamknięty obieg maszyn rozmawiających z maszynami. Mandat człowieka bez rewolucji agentowej to administracja, która nie nadąga za skalą i szybkością spraw.

Wymiar	Praktyka obecna	Formuła naczelna
Przedmiot projektu	pojedynczy system, chatbot lub e-usługa	para sprzężonych narzędzi po obu stronach relacji państwo–obywatel
Rola AI	wsparcie redakcyjne, wyszukiwarka, odpowiedź na pytanie	agent wykonujący zadania w granicach mandatu, pod warstwą kontrolną
Rola obywatela	petent wypełniający formularz	podmiot z własnym agentem, portfelem danych i prawem kontestacji
Przejrzystość	publikacja dokumentu lub rejestru	maszynowo czytelne zobowiązania i narzędzie do ich odczytania
Miara sukcesu	liczba wdrożeń i e-usług	zmiana realnego rozkładu podmiotowości, czas i jakość rozstrzygnięcia

Metryka i status dokumentu

Pozycja	Wartość
Tytuł	Narzędzia AI — dla obywateli i administracji TechnoRepubliki
Wydanie	drugie, przebudowane i poszerzone (wydanie pierwsze: 08.08.2026)
Pozycja w serii	dokument programowy TechnoSanacji nr 44; rozwija ósmy postulat „Administracja AI z człowiekiem w pętli decyzyjnej”
Formuła naczelna	symetryczna i sprzężona architektura narzędzi AI obywatela i administracji pod mandatem człowieka
Podstawa badawcza	destylat NAI: 1610 publikacji w pięciu klastrach (GOVTECH 303, MCDM 159, CITIZEN 304, ETHICS 492, CITIZENTOOLS 352), 54 tezy częściowe
Podstawa programowa	Manifest, Projekt Konstytucji, Administracja państwowa, Controlling cybernetyczny, System informacyjny, Graf multi-agentów, Rezyliencja, Program 2G
Uzasadnienia	cybernetyczne, systemowe, sieciowe i technologiczne (rozdział 2)
Status	dokument programowy do debaty publicznej; propozycje ustrojowe wymagają legalnej ścieżki zmiany prawa i Konstytucji
Licencja	Creative Commons — Uznanie autorstwa
Data	16 września 2026 r.

Dwanaście zobowiązań symetrycznej architektury narzędzi AI

Zobowiązania są operacyjnym rozwinięciem formuły naczelnej. Mechanizm, którego nie da się przypisać do żadnego z nich, nie wchodzi do programu.

1. Żadne narzędzie AI administracji nie jest wdrażane bez wskazania narzędzia obywatela, które pozwala je zrozumieć, wykorzystać lub zakwestionować.
2. Każdy agent działający w imieniu państwa ma tożsamość maszynową, mandat, właściciela, limity uprawnień i przechodzi przez warstwę kontrolną przed wykonaniem czynności.
3. Każdy obywatel ma prawo używać własnego agenta w kontaktach z państwem, a państwo ma obowiązek udostępnić mu maszynowo czytelny interfejs.
4. Tożsamość i unikalność człowieka są potwierdzane kryptograficznie i z ochroną prywatności — nigdy przez centralną megabazę.
5. Dane obywatela są przechowywane i udostępniane w modelu portfela poświadczeń; państwo nie żąda danych, które już posiada.
6. Decyzje prawnie doniosłe są rozstrzygane regułami jawnymi (rules-first); modele stochastyczne mogą je wspierać, lecz nie rozstrzygać.
7. Nadzór człowieka jest stopniowany według ryzyka i ma realną pojemność: czas, kompetencje, prawo zatrzymania i symetrię uzasadnienia.
8. Rejestr algorytmów publicznych jest obowiązkowy, ustrukturyzowany, wersjonowany i dostępny przez API dla narzędzi obywatelskich.
9. Wartości konstytucyjne są zakotwiczone w systemie jako nadrzędne ograniczenie agentów i ponownie weryfikowane w każdym cyklu.
10. Infrastruktura narzędzi obywatelskich jest publiczna i dostępna dla wszystkich; nie może zależeć od siły nabywczej użytkownika.
11. Każde wdrożenie agenta obywatelskiego jest połączone z reformą pojemności usług, aby automatyzacja wniosków nie obróciła się przeciw najślabszym.
12. Wynik mierzy się zmianą realnej podmiotowości i jakości rozstrzygnięć, a każdy miernik ma wskaźnik ryzyka wykrywający jego nadużycie.

SPIS TREŚCI

Formuła naczelna	2
Metryka i status dokumentu	3
Dwanaście zobowiązań symetrycznej architektury narzędzi AI	4
Status dokumentu i spójność strategiczna	7
Streszczenie wykonawcze	7
Mapa dokumentu	9
1. Rewolucja agentowa: dlaczego potrzebna jest nowa formuła	9
1.1. Od formuły służby do formuły symetrii	9
1.2. Pięć przełomów technologicznych lat 2024–2026	10
1.3. Trzy skutki dla relacji państwo–obywatel	11
1.4. Kalendarz regulacyjny 2026–2028	11
2. Formuła naczelna, dwanaście tez i cztery drogi dowodowe	12
2.1. Dwanaście tez naczelnych publikacji	12
2.2. Uzasadnienie cybernetyczne	14
2.3. Uzasadnienie systemowe	16
2.4. Uzasadnienie sieciowe	17
2.5. Uzasadnienie technologiczne	18
2.6. Test czterech uzasadnień	19
3. Architektura symetryczna: cztery warstwy	19
3.1. Warstwa zerowa: tożsamość i unikalność człowieka	20
3.2. Warstwa infrastrukturalna: DPI	21
3.3. Warstwa operacyjna: agenci obu stron	21
3.4. Warstwa kontrolna i instytucjonalna	21
3.5. Macierz par sprzężonych	21
4. Mapa zadań, poziomy autonomii i stopniowany nadzór	23
4.1. Mapa zadań możliwych do automatyzacji	23
4.2. Poziomy autonomii A0–A5	24
4.3. Macierz stopniowanego nadzoru: HIC, HITL, HOTL	24
4.4. Warunki znaczącego nadzoru	25
4.5. Dyrektywa wyprzedzająca dla decyzji poza czasem ludzkiej percepcji	25
5. Narzędzia AI administracji	26
5.1. Asystent urzędnika	26
5.2. Warstwa kontrolna agentów (Organizational Control Layer)	26
5.3. Elektroniczne zarządzanie dokumentacją, Narodowy Graf Wiedzy i RAG	27
5.4. AI w legislacji, podatkach, zamówieniach i audycie	29
5.5. Decyzje wielokryterialne: AHP jako tłumacz ocen modeli	29
5.6. Systemy wsparcia decyzji z zachowaniem autonomii decydenta	30
5.7. AIOps i zarządzanie infrastrukturą państwa	30
6. Narzędzia AI obywatela	31
6.1. Jedna brama usług i zdarzenia życiowe	31
6.2. Agent obywatelski i problem zalewu wniosków	32
6.3. Portfel danych i samosuverenna tożsamość	33
6.4. Samoobrona prawna i deterministyczny walidator	33
6.5. Obywatelski audyt algorytmów	34
6.6. Prosty język i dostępność jako infrastruktura	34
6.7. Partycypacja, synteza konsensusu i ochrona konsultacji	35
6.8. Sygnaliści i kanał zwrotny z obowiązkiem reakcji	35

6.9. Asystent dla osób wykluczonych cyfrowo	35
7. Prawa obywatela, algorytmiczny due process i Public Constitutional AI	36
7.1. Prawo do dobrej administracji w epoce agentów	36
7.2. Wyjaśnienia dla trzech odbiorców	37
7.3. Public Constitutional AI	37
7.4. Człowiek w pętli jako obowiązek prawny	37
8. Zarządzanie modelami, agentami i danymi	38
8.1. Rejestr algorytmów przeciw teatrowi przejrzystości	39
9. Bezpieczeństwo, odporność i suwerenność	39
9.1. Nowe klasy ryzyka agentów	39
9.2. Suwerenność jako zarządzana współzależność	41
10. Gotowość organizacyjna, kadry i Program 2G	41
10.1. Poziomy gotowości	41
10.2. Kadry i nowe role	42
10.3. Program 2G: tempo i pamięć	43
11. Pilotáže, KPI, KRI i rozliczalność	43
11.1. Bramy wdrożeniowe	43
11.2. Pulpit wskaźników	44
12. Mapa drogowa 2026–2035	45
13. Test zgodności z formułą, napięcia i warunek falsyfikacji	47
13.1. Test zgodności rozdziałów z formułą naczelną	47
13.2. Napięcia nierozstrzygnięte	47
13.3. Warunek falsyfikacji	48
13.4. Pięć warunków, przy których architekturę należałoby zmienić	48
REKOMENDACJE KOŃCOWE — 12 DECYZJI	48
ANEKS A. MACIERZ KLASYFIKACJI ZASTOSOWANIA AI	49
ANEKS B. MINIMALNA KARTA AGENTA / SYSTEMU AI	49
ANEKS C. KWIT DECYZJI AI	50
ANEKS D. KARTA PILOTAŻU I RAPORT PUBLICZNY	50
ANEKS E. KARTA PARY SPRZĘŻONEJ	51
ANEKS F. SŁOWNIK POJĘĆ, TERMINÓW I SKRÓTÓW	51
F.1. Skróty, akronimy i oznaczenia	51
F.2. Pojęcia sztucznej inteligencji i systemów agentowych	53
F.3. Pojęcia cybernetyczne	55
F.4. Pojęcia systemowe i sieciowe	55
F.5. Pojęcia prawne, instytucjonalne i programowe	56
ANEKS G. BIBLIOGRAFIA	57
G.1. Dokumenty programowe TechnoSanacji	57
G.2. Akty prawne, orzeczenia i dokumenty publiczne	57
G.3. Cybernetyka, teoria systemów i teoria sieci	58
G.4. Sztuczna inteligencja, agenci i governance	58
G.5. Administracja publiczna, GovTech i obywatel	59
G.6. Korpus źródłowy destylatu NAI	60
NARZĘDZIA, KTÓRE SŁUŻĄ — NIE SYSTEM, KTÓRY DOMINUJE	60

Status dokumentu i spójność strategiczna

Dokument rozwija ósmy postulat TechnoSanacji: „Administracja AI z człowiekiem w pętli decyzyjnej”. Łączy go z postulatem jednolitej egzekutywy prezydenckiej, wspólnot lokalnych zamiast rozbudowanych urzędów samorządowych, konkursowego doboru kadr oraz cybernetycznego controllingu wyniku. Nie jest opisem jednego programu informatycznego. Jest projektem architektury, w której narzędzia AI państwa i narzędzia AI obywatela powstają jako jeden sprzężony system.

Pierwsze wydanie dokumentu koncentrowało się na administracji: jej procesach, danych, modelach, agentach i kontroli człowieka. Wydanie drugie zachowuje ten dorobek, lecz zmienia punkt ciężkości. Po pierwsze, uwzględnia przełom agentowy lat 2024–2026, który przesunął AI z roli asystenta do roli wykonawcy. Po drugie, wprowadza drugą, równorzędną stronę relacji: obywatela wyposażonego we własne narzędzia AI. Po trzecie, opiera tezy na destylacie 1610 publikacji naukowych i raportów z pięciu klastrów badawczych oraz poddaje je czterem niezależnym drogom dowodowym.

Koncepcja zakłada, że administracja państwowa pozostaje jednolitym aparatem wykonawczym podporządkowanym konstytucyjnej egzekutywie, natomiast wspólnoty lokalne odzyskują przestrzeń do samoorganizacji, przedsiębiorczości, spółdzielczości, wolontariatu i inicjatyw obywatelskich. W obecnym porządku prawnym samorząd terytorialny ma umocowanie konstytucyjne, a gmina wykonuje zadania publiczne we własnym imieniu i na własną odpowiedzialność. Model opisany w tym dokumencie jest więc propozycją ustrojową wymagającą legalnej ścieżki: debaty, ustaw przejściowych, zmiany Konstytucji i mandatu Narodu w referendum. Ciągłość usług, prawa pracowników oraz ochrona mieszkańców muszą być zachowane na każdym etapie transformacji.

TEZA CENTRALNA

Narzędzie AI po jednej stronie relacji — obywatela albo administracji — bez symetrycznego narzędzia po drugiej stronie nie tworzy podmiotowości, lecz nową fasadę. Agent obywatelski bez maszynowo czytelnych zobowiązań państwa zgaduje interfejs. Rejestr algorytmów bez narzędzia obywatelskiego do jego odczytania jest teatrem przejrzystości. AI-doradca państwa bez zakotwiczonych wartości dryfuje do technokratycznej konwencji. Synchronizacja obu stron jest warunkiem, nie dodatkiem.

Streszczenie wykonawcze

1. AI przeszła od odpowiadania do działania. Agenci planują i wykonują wieloetapowe zadania, a długość zadań, które potrafią ukończyć, rośnie wykładniczo. Państwo zaprojektowane wyłącznie dla ludzkiego wykonawcy traci sterowność.
2. Teza centralna publikacji: narzędzie AI po jednej stronie relacji państwo–obywatel bez symetrycznego narzędzia po drugiej stronie tworzy fasadę, a nie podmiotowość. W każdym z pięciu klastrów badawczych ten wzorzec powraca niezależnie.
3. Architektura symetryczna ma cztery warstwy: zerową (tożsamość i unikalność człowieka), infrastrukturalną (tożsamość, płatności i wymiana danych), operacyjną (agenci obu stron) oraz kontrolną i instytucjonalną.
4. Agent administracji działa pod warstwą kontrolną, która kwalifikuje każdą propozycję czynności: zatwierdź, popraw, zablokuj albo eskaluj. Kontrola obejmuje także komunikację między agentami.

5. Agent obywatela redukuje koszty poznawcze, zgodności i psychologiczne kontaktu z państwem, ale musi być wdrażany razem z reformą pojemności usług i z maszynowo czytelnymi zobowiązaniami administracji.
6. Decyzje prawnie doniosłe są rozstrzygane regułami jawnymi. Nadzór człowieka jest stopniowany według ryzyka i ma mierzalną pojemność; human-in-command jest zdolnością, nie podpisem.
7. Każdy obywatel ma prawo wiedzieć, że użyto AI, otrzymać Kwit Decyzji, kontakt do odpowiedzialnego człowieka i skuteczną drogę odwołania. Legitymizację zasad AI buduje proces współtworzenia, kontestacji i widocznego egzekwowania.
8. Narodowy Graf Wiedzy i RAG korzystają wyłącznie z autorytatywnego, wersjonowanego i bi-temporalnego korpusu prawa. System nie cytuje źródła, którego nie ma w kanonicznym zbiorze.
9. Suwerenność AI jest zarządzana współzależnością w trzech wymiarach: operacyjnym, danych i oprogramowania. Kryterium stanowi kontrola wersji, kluczy i zdolność wyjścia od dostawcy.
10. Wdrożenia są mierzone telemetrycznie z baseline’em i grupą porównawczą; Program 2G łączy Gen Z z seniorami w modelu dual-key, a mapa drogowa prowadzi od inwentaryzacji w 100 dni do jednolitej, suwerennej administracji w latach 2030–2035.

Podstawa: TechnoSanacja — Postulaty i Struktura; destylat NAI „Narzędzia AI dla obywateli i administracji TechnoRepubliki” (1610 publikacji); wydanie pierwsze „Narzędzia AI” (08.08.2026); projekt Konstytucji TechnoRepubliki Polskiej; aktualne prawo RP i UE.



Rysunek 2. Transformacja od urzędu obsługującego formularz do administracji odpowiedzialnej za wynik.

Mapa dokumentu

Część	Zakres	Funkcja
1–2	Rewolucja agentowa i formuła	Przełomy technologiczne, dwanaście tez i cztery drogi dowodowe.
3–4	Architektura i autonomia	Cztery warstwy symetrii, pary sprzężone, mapa zadań i poziomy nadzoru.
5	Narzędzia administracji	Asystent urzędnika, warstwa kontrolna agentów, RAG, decyzje wielokryterialne, AIOps.
6	Narzędzia obywatela	Brama usług, agent obywatelski, portfel danych, samoobrona prawna, audyt, partycypacja.
7–9	Prawa, zarządzanie i bezpieczeństwo	Due process, Public Constitutional AI, rejestr, cykl życia, suwerenność i odporność.
10–12	Zdolność i wdrożenie	Gotowość, kadry 2G, pilotaże, KPI/KRI, mapa drogowa 2026–2035.
13	Weryfikacja	Test zgodności z formułą, napięcia nierozstrzygnięte i warunek falsyfikacji.
A–G	Aneksy	Macierz, Karta Agenta, Kwit Decyzji, karta pilotażu, Karta Pary Sprzężonej, słownik i bibliografia.

1. Rewolucja agentowa: dlaczego potrzebna jest nowa formuła

Od narzędzia, które odpowiada, do agenta, który działa

1.1. Od formuły służby do formuły symetrii

Formuła pierwszego wydania — narzędzia AI realizują algorytm służby — opisywała AI jako instrument w rękach państwa, skierowany ku obywatelowi. Taki model był adekwatny, dopóki AI pozostawała wyszukiwarką, tłumaczem i redaktorem pism. Obywatel był odbiorcą lepszej usługi, a administracja jej wykonawcą. Asymetria między stronami była wtedy ograniczona fizyczną wydajnością ludzi po obu stronach okienka.

Przełom agentowy usuwa to ograniczenie. Agent administracji może w ciągu godziny przejrzeć tysiące spraw, zestawić dane z wielu rejestrów i przygotować rozstrzygnięcia. Agent obywatela może w tym samym czasie złożyć setki pism, odwołań i wniosków o informację. Jeżeli tylko jedna strona dysponuje agentem, powstaje jeden z dwóch układów patologicznych: państwo, które widzi obywatela w całości, podczas gdy obywatel nie rozumie państwa, albo państwo zalewane wnioskami, które odpowiada limitami i opłatami uderzającymi w najłabszych. Nowa formuła nie zastępuje idei służby. Określa warunek, bez którego służba w epoce agentów zamienia się w dominację albo w paraliż.

CIĄGŁOŚĆ I ZMIANA

Z pierwszego wydania pozostają: mapa automatyzacji według odwracalności, human-in-command jako zdolność, Kwit Decyzji, bi-temporalny Graf Wiedzy, rejestr algorytmów, suwerenność warstwowa i model 2G. Nowe są: symetria narzędzi, warstwa zerowa tożsamości, warstwa kontrolna agentów, narzędzia obywatelskie oraz cztery drogi dowodowe.

1.2. Pięć przełomów technologicznych lat 2024–2026

Radykalność obecnej zmiany nie wynika z jednego wynalazku, lecz z nałożenia się pięciu przełomów, które wzajemnie się wzmacniają. Każdy z nich osobno byłby istotny; razem zmieniają ekonomię i architekturę pracy publicznej.

Przełom	Istota zmiany	Skutek dla państwa
Agenci wykonawczy	modele planują zadanie, wywołują narzędzia, obsługują przeglądarkę i systemy, sprawdzają wynik i poprawiają błędy	sprawa może być prowadzona end-to-end przez system; pytanie „kto wykonał czynność” wymaga tożsamości maszynowej
Rozumowanie w czasie wnioskowania	modele przeznaczają więcej obliczeń na analizę przed odpowiedzią, co podnosi trafność w zadaniach wieloetapowych	AI wkracza w obszar analizy prawnej i wariantowej, dotąd zarezerwowany dla specjalistów
Protokoły interoperacyjności agentów	Model Context Protocol (listopad 2024) łączy agentów z narzędziami i danymi; Agent2Agent (kwiecień 2025) standaryzuje komunikację między agentami; oba przekazano pod neutralny zarząd fundacji Linux Foundation	agenci różnych organów i obywateli mogą się porozumiewać; luki bezpieczeństwa protokołów stają się problemem ustrojowym
Załamanie kosztu wnioskowania	koszt zapytania do modelu na poziomie GPT-3.5 spadł między listopadem 2022 a październikiem 2024 ponad 280-krotnie (Stanford AI Index 2025)	masowe użycie agentów jest ekonomicznie dostępne dla każdego obywatela, nie tylko dla państwa i korporacji
Wydłużanie horyzontu zadań	długość zadań, które agenci kończą z 50-procentową skutecznością, podwajała się w latach 2019–2025 mniej więcej co siedem miesięcy, a w ostatnich pomiarach szybciej (METR)	zakres pracy delegowalnej na agenta rośnie szybciej niż cykl legislacyjny i budżetowy państwa

Do przełomów globalnych dochodzi warstwa krajowa. Polska dysponuje otwartymi modelami językowymi rozwijanymi dla języka polskiego i administracji (rodzina PLLuM, model Bielik), krajowymi centrami obliczeniowymi oraz aplikacją mObywatel, do której ma zostać włączony Europejski Portfel Tożsamości Cyfrowej. To aktywa, które pozwalają budować symetrię na własnej infrastrukturze, a nie wyłącznie na usługach zewnętrznych dostawców.

REWOLUCJA AGENTOWA 2024-2026: PIĘĆ PRZEŁOMÓW

1. AGENCI WYKONAWCZY	2. ROZUMOWANIE	3. PROTOKOŁY MCP · A2A	4. KOSZT WNIOSKOWANIA	5. HORYZONT ZADAŃ
AI działa, nie tylko odpowiada planuje, używa narzędzi, obsługuje systemy, poprawia błędy	obliczenia w czasie wnioskowania wyższa trafność w analizie wieloetapowej	agenci łączą się ze sobą standard narzędzi i danych; komunikacja agent-agent	ponad 280× taniej poziom GPT-3.5: XI 2022 → X 2024 (AI Index 2025)	podwojenie ok. co 7 miesięcy w latach 2019–2025, ostatnio szybciej (METR)
Skutek dla państwa sprawa end-to-end; potrzebna tożsamość maszynowa	Skutek dla państwa AI wchodzi w analizę prawną i wariantową	Skutek dla państwa interoperacyjność i nowe luki łańcucha delegacji	Skutek dla państwa agent dostępny dla każdego obywatela	Skutek dla państwa zakres delegacji rośnie szybciej niż cykl prawa

Przełomy wzmacniają się nawzajem: tani, rozumujący agent połączony protokołem z usługami zmienia rozkład władzy między państwem a obywatelem.

Rysunek 3. Pięć przełomów rewolucji agentowej i ich skutki dla relacji państwo–obywatel.

1.3. Trzy skutki dla relacji państwo–obywatel

Asymetria prędkości

Agent pracuje w czasie maszynowym, człowiek w czasie biologicznym. Gdy administracja wykorzystuje agentów, a obywatel nie, różnica prędkości staje się różnicą władzy: państwo zestawia dane i rozstrzyga szybciej, niż obywatel zdąży zrozumieć, jakie dane zostały użyte. Gdy obywatel wykorzystuje agenta, a administracja nie, różnica działa odwrotnie. Badania zebrane w destylacie dokumentują 84 przypadki przeciążenia usług publicznych przez wnioski generowane automatycznie w 11 krajach (zjawisko agentic flooding).

Erozja weryfikowalności

Generatywne modele obniżyły koszt wytworzenia wiarygodnie wyglądającego tekstu, głosu, obrazu i tożsamości niemal do zera. Konsultacje społeczne, petycje, skargi i komentarze do projektów ustaw stają się podatne na farmy botów i kampanie astroturfingowe. Bez kryptograficznego potwierdzenia, że za daną aktywnością stoi unikalny człowiek, każde narzędzie partycypacji jest podatne na atak skalą.

Przesunięcie odpowiedzialności

Im więcej czynności wykonuje agent, tym łatwiej rozmyć odpowiedzialność między organ, dostawcę modelu, operatora i integratora. Skala agenta przekracza fizyczną zdolność ludzkiej weryfikacji każdej sprawy, co rodzi iluzję nadzoru: człowiek formalnie obecny, realnie nieobecny. Rozwiązaniem nie jest zatwierdzanie każdej czynności, lecz stopniowany nadzór z człowiekiem jako arbitrem ryzyka w kluczowych punktach.

1.4. Kalendarz regulacyjny 2026–2028

Rewolucja technologiczna zbiega się w czasie z wejściem w fazę stosowania europejskich i polskich ram prawnych. Harmonogram jest istotny, ponieważ wyznacza okno, w którym Polska może zaprojektować architekturę symetryczną zamiast dopasowywać się do rozwiązań narzuconych przez dostawców.

Data	Zdarzenie
22.01.2026	Singapur ogłasza pierwszy na świecie rządowy Model AI Governance Framework for Agentic AI — dedykowane ramy dla systemów autonomicznych.
maj 2026	Porozumienie polityczne w sprawie pakietu Digital Omnibus on AI: obowiązki dla systemów wysokiego ryzyka z Załącznika III AI Act przesunięte na 2.12.2027, z Załącznika I na 2.08.2028 (wiążące po publikacji w Dzienniku Urzędowym UE).
24.07.2026	Prezydent RP podpisuje ustawę o systemach sztucznej inteligencji; ustawa ogłoszona 27.07.2026, zasadnicza część przepisów obowiązuje od 11.08.2026.
2.08.2026	Zaczynają obowiązywać obowiązki przejrzystości z art. 50 AI Act, w tym informowanie osoby, że prowadzi interakcję z systemem AI.
28.10.2026	W Polsce zaczynają obowiązywać przepisy o kontroli, postępowaniach skargowych i karach; skargę do Komisji Rozwoju i Bezpieczeństwa Sztucznej Inteligencji (KRiBSI) może złożyć każda osoba.
koniec 2026	Termin udostępnienia obywatelom europejskich portfeli tożsamości cyfrowej na podstawie eIDAS 2.0; w Polsce pilotaż portfela w aplikacji mObywatel.
2.08.2027	Przesunięty termin utworzenia krajowych piaskownic regulacyjnych AI.
2.12.2027	Stosowanie obowiązków dla samodzielnych systemów AI wysokiego ryzyka (m.in. dostęp do świadczeń i usług publicznych, zatrudnienie, edukacja, migracja).

WNIOSEK Z KALENDARZA

Prawo UE i RP wyznacza minimum ochrony, lecz nie zawiera architektury symetrii. AI Act reguluje dostawców i podmioty stosujące, nie przewiduje jednak prawa obywatela do własnego agenta ani obowiązku państwa do maszynowo czytelnych zobowiązań. Tę lukę ma wypełnić program TechnoRepubliki.

Podstawa: destylat NAI, soczewki I–V; Stanford HAI, AI Index Report 2025; METR, Measuring AI Ability to Complete Long Tasks (2025) i Time Horizon 1.1 (2026); IMDA, Model AI Governance Framework for Agentic AI (2026); Gibson Dunn, EU AI Act Omnibus Agreement (27.05.2026); ustawa o systemach sztucznej inteligencji (2026).

2. Formuła naczelną, dwanaście tez i cztery drogi dowodowe

Formuła nie jest mottem, lecz zdaniem sprawdzalnym

2.1. Dwanaście tez naczelných publikacji

Tezy zostały wyprowadzone z destylatu 1610 publikacji podzielonych na pięć soczewek badawczych: architekturę GovTech (I), metody wielokryterialne i cybernetykę decyzyjną (II), ekosystem i interfejs obywatela (III), etykę, prawo i suwerenność (IV) oraz obywatelskie narzędzia AI (V). Każda teza wskazuje konsekwencję operacyjną i miejsce, w którym jest rozwinięta.

Nr	Teza	Konsekwencja operacyjna	Podstawa / rozdział
T1	Przełom agentowy zmienia jakościowo naturę pracy publicznej: AI przestaje być narzędziem, staje się wykonawcą.	projektowanie procesów dla wykonawcy ludzkiego i maszynowego jednocześnie	przełomy 2024–2026; rozdz. 1
T2	Narzędzie AI po jednej stronie relacji bez symetrycznego narzędzia po drugiej tworzy fasadę, nie podmiotowość.	zasada pary sprzężonej jako warunek wdrożenia	teza centralna destylatu; rozdz. 3.5
T3	Tożsamość i unikalność człowieka są warstwą zerową; bez nich każde narzędzie partycypacji jest podatne na atak Sybil.	personhood credential i portfel SSI przed narzędziami partycypacji	V.11, V.3, III.7, II.3; rozdz. 3.1
T4	Interfejs AI jest bezwartościowy bez cyfrowej infrastruktury publicznej pod nim; jeden punkt wejścia to sieć wyspecjalizowanych agentów.	DPI w modelu klepsydry; federacja agentów zamiast monolitu	III.1, III.2, I.10; rozdz. 3.2, 6.1
T5	Agent administracji musi przechodzić przez warstwę kontrolną, która naprawia propozycje, a nie tylko je blokuje; kontrola obejmuje komunikację agent-agent.	Organizational Control Layer z dziennikiem audytowym	I.1, I.2, I.4; rozdz. 5.2
T6	Nadzór człowieka jest stopniowany według ryzyka, a decyzje prawnie doniosłe są rozstrzygane regułami jawnymi.	macierz HIC/HITL/HOTL; rules-first	IV.1–3, V.6; rozdz. 4
T7	Agent obywatelski redukuje obciążenia administracyjne, lecz bez reformy pojemności i maszynowo czytelnych zobowiązań państwa wywołuje zalew wniosków.	standard zobowiązań maszynowych i reforma pojemności usług	V.1, V.12; rozdz. 6.2
T8	Przejrzystość algorytmów wymaga narzędzia odczytu po stronie obywatela; rejestr bez krytycznego odbiorcy jest teatrem.	rejestr ustrukturyzowany z API i obywatelski toolkit audytu	IV.6–8, V.4; rozdz. 6.5, 8
T9	AI-doradca państwa bez zakotwiczonych wartości dryfuje; ocena LLM wymaga warstwy korygującej i ochrony przed manipulacją.	System 5 re-iniekcjonowany co cykl; AHP jako tłumacz ocen	II.1–7; rozdz. 2.2, 5.5
T10	Pełna suwerenność AI jest strukturalnie niewykonalna; realna strategia to zarządzana współzależność w trzech wymiarach.	mapa zależności i reguła najbliższej warstwy	IV.4–5; rozdz. 9
T11	Legitymizację zasad AI tworzy proces: współtworzenie reguł, mechanizm kontestacji i widoczne egzekwowanie.	Public Constitutional AI; agent obywatela doradczy i wykonawczy, nigdy głosujący ustrojowo	V.2, V.10; rozdz. 7
T12	AI w administracji nie może kopiować sektora prywatnego; wymaga reformy organizacyjnej równoległej z wdrożeniem i publicznej infrastruktury przeciw nierówności agentowej.	reforma instytucjonalna przed technologią; Program 2G	IV.12, III.10–11, V.9; rozdz. 10

Tezy T1–T2 opisują naturę zmiany, T3–T4 fundament, T5–T8 sposób działania narzędzi po obu stronach, T9–T12 warunki, bez których architektura obraca się przeciw swojemu celowi.

2.2. Uzasadnienie cybernetyczne

Cybernetyka jest nauką o sterowaniu i komunikacji w systemach złożonych. Dla państwa epoki agentów dostarcza czterech twierdzeń, które przesądzą o konieczności symetrii.

Prawo wymaganej różnorodności

Zgodnie z prawem Ashby’ego regulator może skutecznie sterować systemem tylko wtedy, gdy jego różnorodność — liczba rozróżnialnych stanów i odpowiedzi — jest co najmniej równa różnorodności zakłóceń. Rewolucja agentowa gwałtownie zwiększa różnorodność po obu stronach: liczbę wniosków, wariantów danych, kanałów i prób obejścia reguł. Administracja wyposażona tylko w ludzi przegrywa z różnorodnością generowaną przez agentów. Obywatel wyposażony tylko we własną uwagę przegrywa z różnorodnością państwa działającego przez agentów. Symetria narzędzi jest więc spełnieniem prawa wymaganej różnorodności dla obu regulatorów jednocześnie: państwo reguluje sprawy, a obywatel reguluje państwo.

Twierdzenie o dobrym regulatorze

Conant i Ashby wykazali, że każdy dobry regulator systemu musi być jego modelem. Agent działający w imieniu obywatela musi mieć model zobowiązań państwa; jeśli ich nie zna, zgaduje interfejs, co w praktyce oznacza kruche pobieranie treści ze stron internetowych. Państwo musi mieć model obywatela, lecz model ten nie może być ukrytym profilem. Maszynowo czytelne zobowiązania administracji i portfel poświadczeń obywatela są dwoma jawnymi modelami, które pozwalają obu regulatorom działać bez wzajemnej inwigilacji.

Model systemu żywotnego

Model systemu żywotnego Stafforda Beera (VSM) rozróżnia pięć funkcji niezbędnych do przetrwania organizacji. Destylat wskazuje, że luki w administracji publicznej dają się precyzyjnie zlokalizować w tym modelu i przełożyć na specyfikację funkcjonalną narzędzi AI.

Funkcja VSM	Rola w państwie	Narzędzie AI administracji	Narzędzie AI obywatela (para)
S1 — operacje	prowadzenie spraw, świadczenie usług	agenci dziedzinowi pod warstwą kontrolną	agent obywatelski, brama usług
S2 — koordynacja	tłumienie oscylacji między jednostkami	telemetria w czasie rzeczywistym, wspólne formaty, once-only	status sprawy, portfel poświadczeń, once-only
S3 — kontrola i S3* audyt	alokacja zasobów, niezależna weryfikacja	automatyczne alerty, KPI/KRI, audyt algorytmów	obywatelski audyt algorytmów, skarga do KRIBSI
S4 — rozpoznanie przyszłości	foresight, adaptacja	pamięć instytucjonalna GenAI, RAD, analiza skutków regulacji	synteza konsensusu, konsultacje odporne na boty
S5 — tożsamość i wartości	zasady nadrzędne, polityka	zakotwiczone wartości konstytucyjne, reguły twarde	Public Constitutional AI: współtworzenie i kontestacja reguł



Rysunek 4. Pętle sterowania narzędzi AI w modelu systemu żywotnego: funkcje S1–S5 po stronie państwa i obywatela.

Kanał algedoniczny i sprzężenie zwrotne

W VSM kanał algedoniczny przenosi sygnał alarmowy z poziomu operacji bezpośrednio do poziomu decyzji, z pominięciem hierarchii. W państwie tę funkcję pełni skarga, sygnalizacja nieprawidłowości i odwołanie. Destylat pokazuje jednak, że samo zbieranie sygnału nie zamyka pętli: kanał zwrotny działa wyłącznie wtedy, gdy po drugiej stronie istnieje instytucja z obowiązkiem i zdolnością reakcji. Narzędzie AI obywatela, które generuje skargę, musi być sprzężone z narzędziem AI administracji, które ją klasyfikuje, przypisuje właścicielowi i monitoruje termin — oraz z organem, który ponosi skutki bezczynności.

Cybernetyka drugiego rzędu i dryf wartości

Cybernetyka drugiego rzędu przypomina, że obserwator jest częścią systemu. Symulacja projektu Cybersyn przeprowadzona z użyciem modeli językowych wykazała, że AI-doradca państwa bez zakotwiczonej funkcji S5 dryfuje z czasem ku konwencjonalnemu, technokratycznemu językowi — niezależnie od radykalności instrukcji początkowej. Wniosek jest konstrukcyjny: wartości konstytucyjne muszą być ponownie wprowadzane do systemu w każdym cyklu decyzyjnym i sprawdzane testem regresji, a nie zapisane raz w dokumencie wdrożeniowym.

WNIOSEK CYBERNETYCZNY

Państwo i obywatel są wzajemnymi regulatorami. Rewolucja agentowa zwielokrotnia różnorodność po obu stronach, więc sterowność zachowa tylko układ, w którym obie strony mają narzędzia o porównywalnej różnorodności, jawne modele siebie nawzajem i zamknięte pętle sprzężenia zwrotnego.

2.3. Uzasadnienie systemowe

Analiza systemowa pyta, gdzie przebiega granica systemu, jakie są jego sprzężenia i w których punktach niewielka zmiana wywołuje trwały skutek.

Granica systemu przesądza o wyniku

Jeśli granicą projektu jest urząd, sukcesem jest szybsze rozpatrzenie wniosku. Jeśli granicą jest relacja państwo–obywatel, sukcesem jest krótsza droga od potrzeby do rozstrzygnięcia i od błędu do naprawy. Destylat pokazuje, że one-stop-shop jest projektem instytucjonalnym, nie technicznym: warunki sukcesu to kolejno mandat polityczny, podstawa prawna, współpraca międzyinstytucjonalna, a dopiero potem technologia. Zasada once-only przynosi policzalny zwrot — w Estonii szacowany na 844–1407 lat pracy rocznie i około 2% PKB — lecz tylko wtedy, gdy granica systemu obejmuje wszystkie rejestry.

Punkty dźwigni

Donella Meadows wskazała, że najsilniejsze punkty dźwigni leżą w regułach, strukturze przepływów informacji i celach systemu, a nie w parametrach. Zakup kolejnego modelu jest zmianą parametru. Obowiązek udostępnienia maszynowo czytelnego interfejsu dla agenta obywatela jest zmianą struktury przepływu informacji. Zasada, że żadne narzędzie administracji nie wchodzi do produkcji bez pary po stronie obywatela, jest zmianą reguły — i dlatego działa trwalej niż jednorazowy program cyfryzacji.

Prawo Goodharta i skutki niezamierzone

Gdy miara staje się celem, przestaje być dobrą miarą. Liczba wdrożonych chatbotów, liczba e-usług i liczba przeszkolonych urzędników są miarami łatwymi do osiągnięcia bez zmiany rozkładu podmiotowości. Dlatego każdy KPI w tej publikacji ma przypisany KRI wykrywający nadużycie. Przypadek australijskiego systemu Robodebt — automatycznego naliczania rzekomych nadpłat świadczeń, zakończony ugodą na 1,2 mld AUD i komisją królewską — pokazuje, że optymalizacja wskaźnika odzysku długu przy braku reguł jawnych i nadzoru prowadzi do masowej krzywdy.

Awarie normalne i sprzężenia ścisłe

Charles Perrow wykazał, że w systemach złożonych o ścisłych sprzężeniach awarie są normalne, a nie wyjątkowe. Łańcuch agentów przekazujących sobie zadania jest systemem ściśle sprzężonym: błąd jednego ogniwa propaguje się szybciej, niż człowiek zdoła zareagować. Wymaga to limitów delegacji, wyłączników obwodu, trybu degradacji i procedury ręcznej dla każdej funkcji zautomatyzowanej.

WNIOSEK SYSTEMOWY

Granicą systemu narzędzi AI jest relacja państwo–obywatel, nie urząd. Najsilniejszą dźwignią jest reguła pary sprzężonej i obowiązek zobowiązań maszynowych. Każda miara wyniku potrzebuje miary nadużycia, a każdy łańcuch agentów — bezpiecznika.

2.4. Uzasadnienie sieciowe

Państwo epoki agentów jest siecią węzłów ludzkich i maszynowych. Teoria sieci pozwala przewidzieć, gdzie powstanie wartość, gdzie koncentracja władzy i gdzie kaskada awarii.

Model klepsydry i efekt sieciowy standardu

Bank Światowy opisuje cyfrową infrastrukturę publiczną jako klepsydrę: wąska szyjka wspólnych standardów tożsamości, płatności i wymiany danych podtrzymuje nieograniczoną liczbę aplikacji po obu stronach. Wartość sieci rośnie z liczbą uczestników korzystających ze wspólnego standardu. Jeśli standard ustala dostawca, efekt sieciowy staje się mechanizmem uzależnienia; jeśli ustala go państwo w sposób otwarty, staje się dobrem wspólnym, z którego korzystają jednocześnie agenci administracji, agenci obywateli i przedsiębiorcy.

Federacja zamiast monolitu

Estoński Bürokratt realizuje ideę „jednego agenta” jako rozproszoną sieć wyspecjalizowanych asystentów za jednym punktem wejścia. Ukraińska Diia zunifikowała ponad dwa tysiące usług pod jedną marką. Sieć federacyjna jest odporniejsza od monolitu, pozwala na specjalizację dziedzinową i ogranicza skutki kompromitacji pojedynczego węzła.

Sieci bezskalowe, huby i atak Sybil

Sieci rzeczywiste mają strukturę bezskalową: nieliczne huby skupiają większość połączeń. Są odporne na przypadkowe awarie, lecz kruche wobec ataku na huby. W sieci narzędzi AI hubami są dostawcy modeli, dostawcy tożsamości i platformy chmurowe — stąd reguła, że żadna funkcja krytyczna nie może zależeć od jednego dostawcy. Z kolei atak Sybil polega na tworzeniu wielu fałszywych tożsamości w celu uzyskania nieproporcjonalnego wpływu. Każdy mechanizm „jedna osoba — jeden głos, jeden komentarz, jeden bot” bez kryptograficznego dowodu unikalności człowieka jest wobec niego bezbronny.

Łańcuchy delegacji i kaskady w sieciach współzależnych

Analiza protokołów agent-agent wykazuje luki nawet przy pełnej zgodności ze specyfikacją: kontekst bez powiązania z właścicielem, tożsamość niepropagowana przez łańcuch delegacji i samodeklarowane zdolności bez atestacji. W sieciach współzależnych awaria w jednej warstwie wywołuje kaskadę w drugiej. Gdy agenci administracji zależą od tej samej tożsamości maszynowej, tej samej chmury i tego samego modelu, jedna kompromitacja może objąć cały łańcuch usług. Warstwa kontrolna musi więc obejmować komunikację między agentami, a tożsamość i mandat muszą być przenoszone przez każdy krok delegacji.

WNIOSEK SIECIOWY

Wartość narzędzi AI tworzy otwarty standard w szyjce klepsydry, a nie liczba aplikacji. Odporność tworzy federacja, dywersyfikacja hubów i kryptograficzna unikalność człowieka. Bezpieczeństwo tworzy tożsamość i mandat przenoszone przez cały łańcuch delegacji.

2.5. Uzasadnienie technologiczne

Czwarta droga dowodowa pyta, czy architektura symetryczna jest wykonalna przy obecnym stanie techniki i jakie ograniczenia technologiczne przesądzą o jej kształcie.

Stochastyczność modeli a rządy prawa

Modele językowe są systemami probabilistycznymi: dla tych samych danych mogą dawać różne wyniki. Rządy prawa wymagają, aby tożsame sprawy były rozstrzygane tożsamo. Destylat wskazuje, że ramy zaufania dla AI w administracji wymuszają zasadę rules-first: reguła jawna i deterministyczna rozstrzyga, model wspiera analizę, wyszukuje, streszcza i proponuje. Tam, gdzie model generuje treść o skutkach prawnych, niezbędny jest deterministyczny walidator — na przykład weryfikator cytowań sprawdzający każde odwołanie w kanonicznym korpusie.

Warstwa kontrolna jako komponent, nie procedura

W badaniu warstwy kontrolnej nad agentami organizacyjnymi przechwytywanie propozycji czynności i kwalifikowanie ich do zatwierdzenia, poprawy, blokady lub eskalacji zredukowało odsetek niebezpiecznych wykonań z 88% do zera. Kluczowy wynik jest subtelniejszy: samo blokowanie przywracało jedynie około 60% skuteczności zadania. Pełną skuteczność przy bezpieczeństwie dawała dopiero ustrukturyzowana naprawa propozycji. Technologicznie oznacza to, że bezpieczeństwo agentów trzeba projektować jako aktywny komponent wykonawczy, a nie jako listę zakazów.

Formalne typowanie interfejsów narzędzi

Testy agentów w zarządzaniu kryzysowym pokazały, że dominującym źródłem błędów nie jest wybór złego narzędzia, lecz niedopasowanie typów parametrów. Interfejsy usług publicznych udostępniane agentom muszą mieć formalny schemat, walidację i jednoznaczną semantykę. To ten sam warunek, który po stronie obywatela nazywamy maszynowo czytelnymi zobowiązaniami państwa.

Kryptografia tożsamości i prywatności

Weryfikowalne poświadczenia, portfele samosuverennej tożsamości, selektywne ujawnianie atrybutów i dowody o wiedzy zerowej pozwalają potwierdzić, że osoba jest pełnoletnia, uprawniona lub unikalna, bez ujawniania pełnych danych. Rozporządzenie eIDAS 2.0 tworzy dla tych technik europejski standard. Polski dowód cyfrowy w mObywatelu, indyjski DigiLocker i brytyjski program Digital ID potwierdzają kierunek: portfel poświadczeń trzymany przez obywatela, nigdy centralna baza rządowa.

Bi-temporalność, proveniencja i zero trust

Prawo zmienia się w czasie, a wyszukiwanie semantyczne nie odróżnia przepisu obowiązującego od uchylonego. Bi-temporalny model danych — czas obowiązywania w świecie i czas wiedzy systemu — pozwala odtworzyć stan prawny na dzień decyzji. Proveniencja i podpisy chronią korpus przed zatruciem, a architektura zero trust wymaga, aby każdy agent, narzędzie i źródło były uwierzytelniane i ograniczane niezależnie od tego, czy działają „wewnątrz urzędu”.

WNIOSEK TECHNOLOGICZNY

Architektura symetryczna jest wykonalna już dziś, ponieważ jej elementy — protokoły agentowe, poświadczenia weryfikowalne, portfele tożsamości, RAG z walidacją i warstwy kontrolne — istnieją i są standaryzowane. Brakuje nie technologii, lecz reguły, która łączy je w pary sprzężone pod mandatem człowieka.

2.6. Test czterech uzasadnień

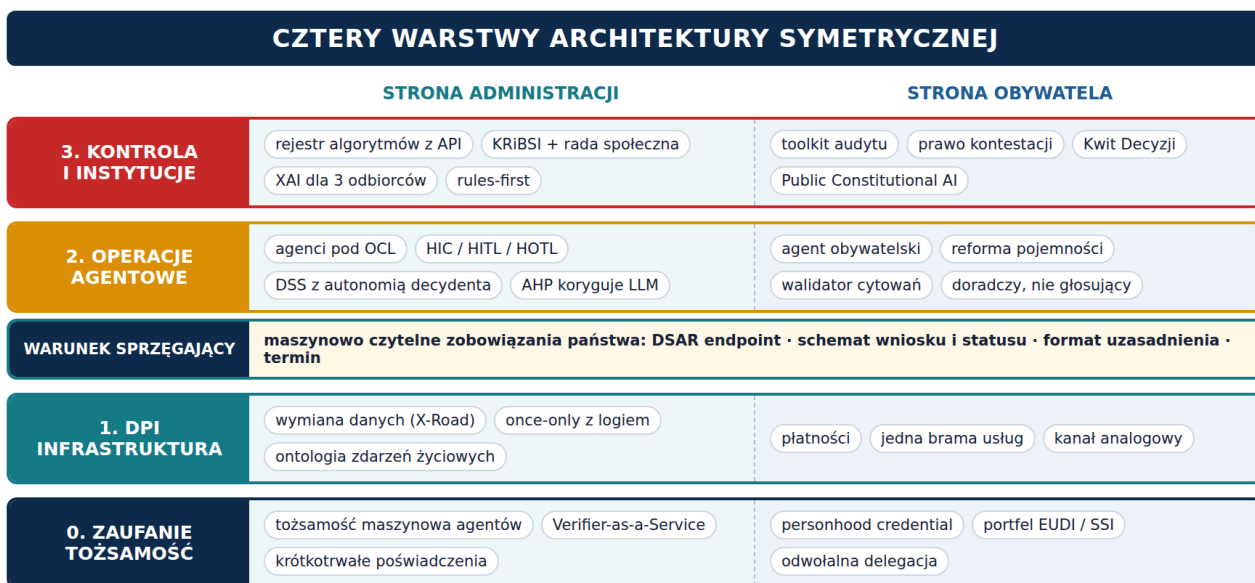
Każdy zasadniczy mechanizm publikacji musi mieć uzasadnienie we wszystkich czterech porządkach. Mechanizm, któremu brakuje choć jednej rubryki, jest traktowany jako hipoteza wymagająca dalszej pracy.

Mechanizm	Cybernetyka	System	Sieć	Technologia
Para sprzężona narzędzi	równość różnorodności regulatorów	granica = relacja, nie urząd	dwustronny efekt sieciowy	wspólne protokoły i schematy
Personhood credential i portfel SSI	wiarygodny sygnał wejściowy	warunek wszystkich usług	odporność na atak Sybil	poświadczenia weryfikowalne, ZKP
DPI w modelu klepsydry	kanal S2 koordynacji	punkt dźwigni struktury	standard jako dobro wspólne	interoperacyjność, once-only
Warstwa kontrolna agentów (OCL)	regulator drugiego stopnia	bezpiecznik sprzężeń ścisłych	kontrola łańcucha delegacji	przechwyt i naprawa propozycji
Maszynowo czytelne zobowiązania	jawny model regulowanego	reguła zamiast parametru	szyjka klepsydry dla agentów	formalnie typowane API
Stopniowany nadzór i rules-first	pojemność regulatora	ochrona przed awarią normalną	ogranicza kaskady	determinizm i walidatory
Rejestr algorytmów z API	kanal S3* audytu	przeciw teatrowi przejrzystości	krytyczny odbiorca w sieci	wersjonowanie, schemat danych
Zakotwiczone wartości S5	przeciw dryfowi drugiego rzędu	cel systemu jako dźwignia	wspólny punkt odniesienia	testy regresji wartości
Kwit Decyzji i kontestacja	kanal algedoniczny	pętla korekty błędu	ścieżka od węzła do organu	proweniencja i log decyzji
Suwerenność trójwymiarowa	autonomia regulatora	reguła najślabszej warstwy	dywersyfikacja hubów	klucze, wersje, prawo wyjścia

3. Architektura symetryczna: cztery warstwy

Najpierw zaufanie i infrastruktura, potem agenci, zawsze kontrola

Pięć soczewek badawczych prowadzi niezależnie do tej samej architektury. Składa się ona z czterech warstw przekrojowych, z których każda obsługuje jednocześnie stronę administracji i stronę obywatela. Kolejność warstw jest kolejnością budowy: warstwa wyższa nie działa bez niższej.



Warstwa wyższa nie działa bez niższej. Bez warstwy zerowej każde narzędzie partycypacji jest podatne na atak Sybil.

Rysunek 5. Cztery warstwy architektury symetrycznej narzędzi AI TechnoRepubliki.

3.1. Warstwa zerowa: tożsamość i unikalność człowieka

Warstwa zerowa odpowiada na pytanie, które w epoce generatywnej AI stało się fundamentalne: czy po drugiej stronie jest człowiek, i czy jest to jeden człowiek. Składają się na nią trzy elementy: poświadczenie osobowości (personhood credential), które kryptograficznie potwierdza unikalność osoby bez ujawniania jej danych; portfel samosuwierennej tożsamości z weryfikowalnymi poświadczeniami, zgodny z europejskim portfelem EUDI; oraz tożsamość maszynowa agentów, powiązana z mandatem i właścicielem.

Element	Po stronie obywatela	Po stronie administracji
Unikalność człowieka	dowód „jestem unikalną osobą” w konsultacjach, petycjach i delegacji do agenta	weryfikacja unikalności bez centralnego rejestru aktywności
Portfel poświadczeń	obywatel trzyma i selektywnie ujawnia atrybuty	organ weryfikuje poświadczenie przez usługę weryfikatora (Verifier-as-a-Service)
Tożsamość agenta	agent obywatela działa na podstawie delegacji z jawnym zakresem	agent organu ma tożsamość maszynową, krótkotrwałe poświadczenia i mandat
Odwołalność	obywatel może cofnąć delegację w każdej chwili	organ może zawiesić agenta i unieważnić jego poświadczenia

Wzorzec Verifier-as-a-Service pozwala jednej warstwie orkestracji obsługiwać weryfikację poświadczeń w wielu formatach przez most OIDC, tak aby urzędy integrowały się konfiguracją, a nie indywidualnym kodem. Bez warstwy zerowej reszta architektury jest podatna na manipulację skalą: farmy botów w konsultacjach, fałszywe petycje, podszywanie się pod agentów organów.

3.2. Warstwa infrastrukturalna: DPI

Cyfrowa infrastruktura publiczna obejmuje tożsamość, płatności i wymianę danych. Jej wzorcami są estoński X-Road, brazylijski Pix i indyjski Aadhaar z ekosystemem India Stack. Bez tej warstwy żaden agent — ani państwowy, ani obywatelski — nie działa ponad silosami resortowymi. Zasada once-only wymaga rozwiązań chroniących prywatność: rozproszonej wymiany danych na żądanie z logiem dostępu widocznym dla obywatela, a nie centralnej megabazy, w której korzyść wydajnościowa rodzi ryzyko nadzoru.

Warstwa infrastrukturalna obejmuje także semantykę. Interfejs trzeba projektować wokół zdarzeń życiowych — narodziny, zmiana pracy, założenie firmy, choroba, śmierć bliskiej osoby — a nie wokół procedur urzędowych. Wymaga to formalnej ontologii rozróżniającej zdarzenie, proces, daną rejestrową i usługę; w przeciwnym razie architektura informatyczna rozjeżdża się między urzędami.

3.3. Warstwa operacyjna: agenci obu stron

Warstwa operacyjna jest miejscem, w którym działają narzędzia opisane w rozdziałach 5 i 6. Po stronie administracji są to agenci dziedzini z warstwą kontrolną, stopniowanym nadzorem, systemami wsparcia decyzji zachowującymi autonomię decydenta oraz metodami wielokryterialnymi korygującymi oceny modeli. Po stronie obywatela są to agent redukujący obciążenia administracyjne, portfel danych, narzędzia samoobrony prawnej, audytu i partycypacji.

WARUNEK SPRZĘGAJĄCY

Obie strony warstwy operacyjnej łączą maszynowo czytelne zobowiązania administracji: ustandaryzowany punkt dostępu do danych osobowych (DSAR), schemat statusu sprawy, format uzasadnienia i termin odpowiedzi. Bez nich agent obywatelski zgaduje interfejs, a administracja nie może odróżnić wniosku uprawnionego od automatycznego zalewu.

3.4. Warstwa kontrolna i instytucjonalna

Warstwa kontrolna zamyka pętlę. Obejmuje obowiązkowy i ustrukturyzowany rejestr algorytmów, narzędzie obywatelskie do jego odczytania jako warunek wstępny, wyjaśnienia zróżnicowane dla trzech odbiorców (obywatela, regulatora, zespołu technicznego), prawo kontestacji konkretnej decyzji, kolegiálny organ nadzoru z radą społeczną oraz zasadę rules-first jako wymóg projektowy. W polskim porządku prawnym punktem wyjścia jest Komisja Rozwoju i Bezpieczeństwa Sztucznej Inteligencji; program TechnoRepubliki proponuje jej uzupełnienie o obywatelską radę społeczną i obowiązek publikacji danych rejestru w formacie maszynowym.

3.5. Macierz par sprzężonych

Macierz jest narzędziem kontrolnym zasady T2. Każdy wiersz wskazuje narzędzie administracji, jego obowiązkową parę po stronie obywatela i ryzyko, które materializuje się, gdy para nie istnieje.

Narzędzie AI administracji	Para po stronie obywatela	Ryzyko bez pary
agent dziedzinowy prowadzący sprawę	agent obywatelski z dostępem do statusu i uzasadnienia	obywatel nie wie, co i dlaczego zrobił system
warstwa kontrolna agentów (OCL)	wgląd w dziennik czynności dotyczących własnej sprawy	kontrola wyłącznie wewnętrzna, bez zewnętrznej rozliczalności
rejestr algorytmów publicznych	obywatelski toolkit rozpoznawania i audytu algorytmów	teatr przejrzystości
once-only i wymiana danych między rejestrami	portfel danych, log dostępu i one-click DSAR	efektywność przekształca się w nadzór
system wsparcia decyzji i scoring ryzyka	Kwit Decyzji z kontrfaktykiem i prawem odwołania	niezaskarżalna władza algorytmiczna
asystent legislacyjny i analiza konsultacji	synteza konsensusu i konsultacje z dowodem unikalności	astroturfing i pozorna partycypacja
automatyczna klasyfikacja skarg	obywatelski generator skargi z walidatorem cytowań	zalew nieskutecznych pism lub ignorowanie zasadnych
AI w kontroli podatkowej i wykrywaniu nadużyć	samoobrona prawna na korpusie spraw o niskim statusie	asymetria procesowa na niekorzyść najślabszych
agenci administracji komunikujący się A2A	delegacja z jawnym zakresem dla agenta obywatela	nieprzypisana odpowiedzialność w łańcuchu
AI-doradca polityczny rządu	Public Constitutional AI: współtworzenie i kontestacja zasad	dryf wartości i technokracja bez mandatu

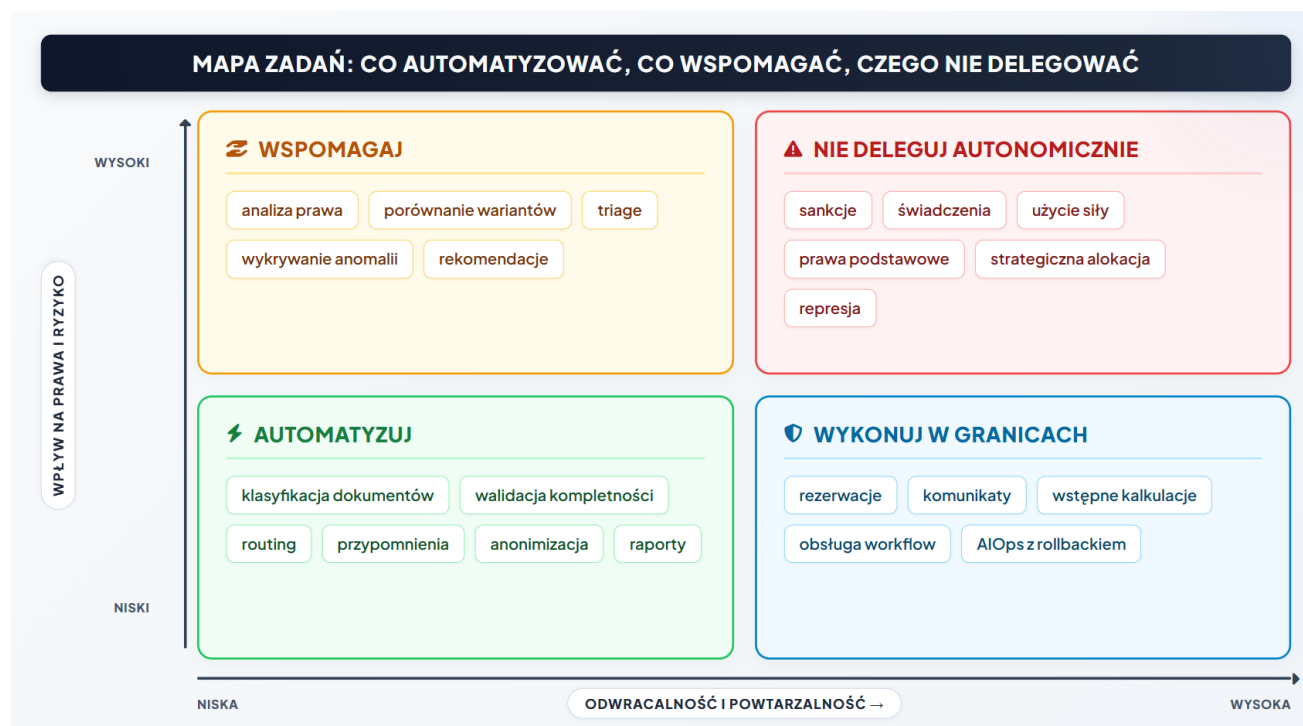
Podstawa: destylat NAI, synteza „pięć soczewek, jedna architektura”; TS-MAS Graf multi-agentów TechnoRepubliki; TS-SI System informacyjny TechnoRepubliki Polskiej; eIDAS 2.0; Bank Światowy — DPI i GovTech Maturity Index.

4. Mapa zadań, poziomy autonomii i stopniowany nadzór

Im większa nieodwracalność i uznaniowość, tym mniejsza autonomia systemu

4.1. Mapa zadań możliwych do automatyzacji

O wyborze zastosowania nie może decydować moda technologiczna. Podstawowym kryterium jest odwracalność skutku, powtarzalność procesu, jakość danych, możliwość obiektywnej walidacji oraz wpływ na prawa obywatela. Przetom agentowy nie zmienia tej zasady — zmienia jedynie to, że zadania o niskim ryzyku mogą być wykonywane end-to-end, a nie tylko wspomagane.



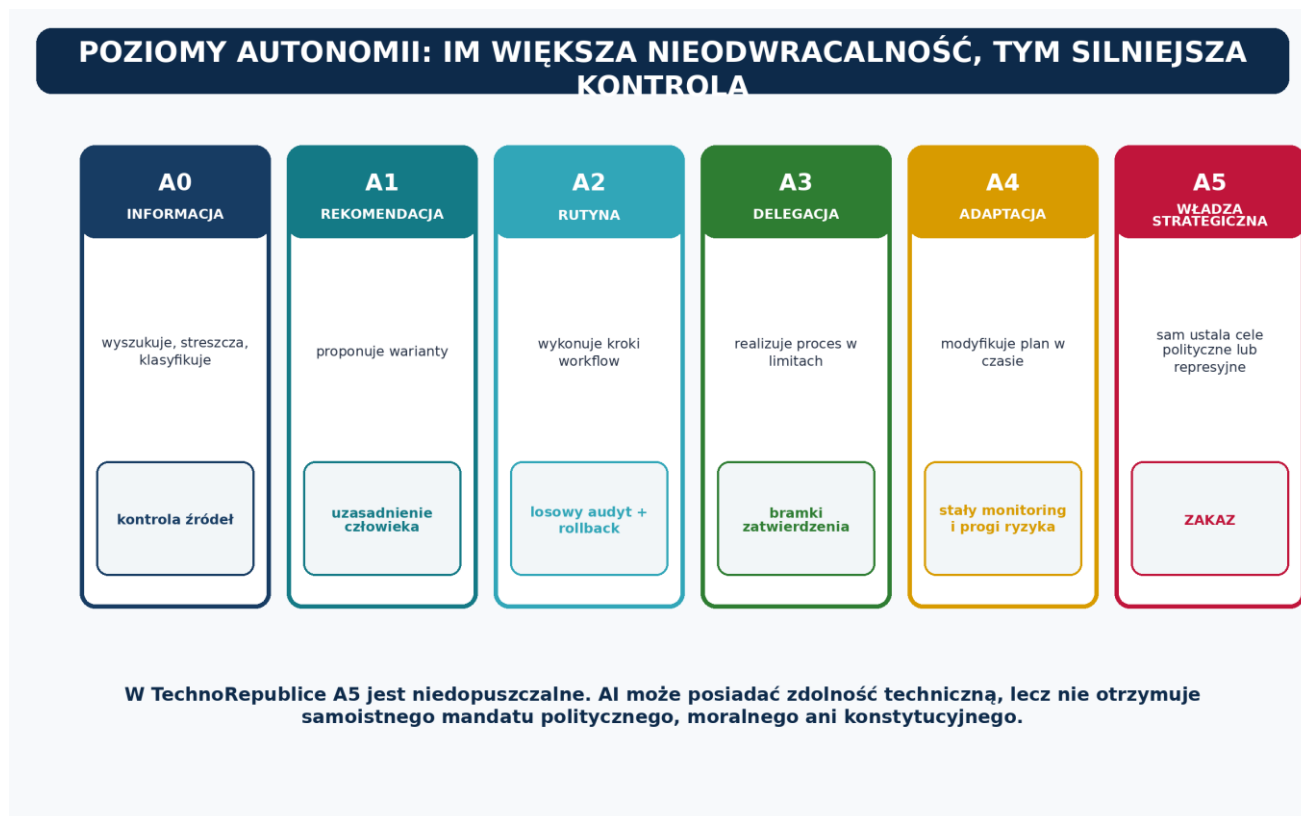
Rysunek 6. Macierz automatyzacji według odwracalności i wpływu na prawa obywatela.

Klasa zadania	Przykłady	Domyślny model
Informacyjne	wyszukiwanie, streszczanie, klasyfikacja, tłumaczenie, proste wyjaśnienie	AI działa; człowiek kontroluje źródła i jakość.
Procesowe odwracalne	routing, przypomnienia, rezerwacje, walidacja kompletności, raporty	automatyzacja agentowa z logiem, SLA i możliwością cofnięcia.
Rekomendacyjne	triale, porównanie wariantów, analiza ryzyka, prognoza	AI proponuje; człowiek uzasadnia decyzję.
Prawnie donioste	świadczenia, sankcje, pozwolenia, wpisy do rejestrów	rules-first, human-in-command, blind review, zasada czterech oczu.
Strategiczne i represyjne	cele polityczne, użycie siły, masowa inwigilacja, represja	zakaz autonomicznej delegacji.

REGUŁA PIERWSZEGO ROKU

Automatyzujemy najpierw czynności powtarzalne, dobrze zdefiniowane i odwracalne. Nie zaczynamy od decyzji o wysokim wpływie na prawa człowieka — i nie wdramy żadnego agenta administracji bez jego pary po stronie obywatela.

4.2. Poziomy autonomii A0–A5



Rysunek 7. Poziomy autonomii i wymagany nadzór w administracji republikańskiej.

Skala A0–A5 opisuje, co system może zrobić bez udziału człowieka: od informacji (A0), przez rekomendację (A1), rutynę odwracalną (A2), delegację procesu w limitach (A3), adaptację planu w czasie (A4), aż po władzę strategiczną (A5), która w TechnoRepublice pozostaje niedopuszczalna. Poziom autonomii jest zapisany w Karcie Agenta (Aneks B) i nie może zostać zwiększony bez przejścia kolejnej bramy wdrożeniowej.

4.3. Macierz stopniowanego nadzoru: HIC, HITL, HOTL

Nadzór człowieka musi być stopniowany według ryzyka. Nadzór „na wszystko” jest kosztowny bez proporcjonalnego zysku i paradoksalnie osłabia uwagę tam, gdzie jest najbardziej potrzebna. Destylat proponuje macierz dwóch osi: wpływu modelu na wynik i konsekwencji błędu.

Wpływ modelu \ Konsekwencja błędu	Niska	Średnia	Wysoka
Niski (informacja, wsparcie)	HOTL — człowiek nad pętlą, monitoring statystyczny	HOTL z próbą losową	HITL — zatwierdzenie w punkcie końcowym
Średni (rekomendacja)	HOTL z progami eskalacji	HITL — zatwierdzenie z uzasadnieniem	HIC — człowiek decyduje, blind review
Wysoki (wykonanie, rozstrzygnięcie)	HITL — zatwierdzenie czynności	HIC — decyzja człowieka, zasada czterech oczu	HIC + rules-first; brak autonomii

Human-in-the-loop ma formalnie różne kształty: monitorowanie, zatwierdzenie w punkcie końcowym i zaangażowaną interakcję. Istnieje przy tym nieunikniony kompromis: im lepsza wyjaśnialność systemu, tym trudniejsza bywa atrybucja odpowiedzialności prawnej, ponieważ człowiek „rozumiejący” wynik łatwiej go przejmuje. Prawo UE tego nie różnicuje, dlatego karta systemu musi wskazywać, który kształt nadzoru obowiązuje.

4.4. Warunki znaczącego nadzoru

Warunek	Wymóg
Limit obciążenia	karta systemu określa maksymalną liczbę spraw na godzinę i na dzień.
Blind review	urzędnik analizuje surowe dane przed zobaczeniem wyniku i rekomendacji modelu.
Symetria uzasadnienia	uzasadnienie jest wymagane zarówno przy przyjęciu, jak i odrzuceniu rekomendacji.
Realne uprawnienie	nadzorca może zatrzymać, cofnąć i eskalować czynność bez kary za spowolnienie.
Certyfikacja	mierzy się indywidualną skuteczność nadzorców i jakość nadpisań.
Zasada czterech oczu	decyzje o wysokim wpływie wymagają niezależnego drugiego przeglądu.
Odwołanie	obywatel ma skuteczny środek prawny i wgląd w podstawę działania systemu.

4.5. Dyrektywa wyprzedzająca dla decyzji poza czasem ludzkiej percepcji

Część czynności agentów — w cyberobronie, w sterowaniu siecią energetyczną, w reagowaniu kryzysowym — zachodzi szybciej, niż człowiek zdola je zauważyć. Destylat proponuje konstrukcję Advance Control Directive, analogiczną do oświadczenia pro futuro: odpowiedzialność przypisuje się momentowi ustanowienia dyrektywy, a nie momentowi wykonania, pod warunkiem pełnej informacji przy jej podpisaniu. Dyrektywa określa klasę sytuacji, dozwolone działania, limity, warunek wygaśnięcia i obowiązek raportu po zdarzeniu.

GRANICA

Decyzje dotyczące wolności, zdrowia, majątku, świadczeń, sankcji i użycia siły nie mogą być autonomicznie podejmowane przez system AI. Dyrektywa wyprzedzająca nie może obejmować tych decyzji wobec konkretnej osoby.

Podstawa: destylat NAI IV.1–3; wydanie pierwsze „Narzędzia AI”, rozdz. 3 i 6; art. 14 i 26 AI Act; TS-CC Controlling cybernetyczny TechnoRepubliki.

5. Narzędzia AI administracji

Urzędnik ma otrzymać lepszą pamięć i bezpiecznego wykonawcę, nie cyfrowego przełożonego

5.1. Asystent urzędnika

Asystent urzędnika skraca czas dotarcia do wiarygodnej informacji, porządkuje dokumenty, wykrywa braki, proponuje warianty i pilnuje terminów. Nie ocenia ukrycie pracownika, nie narzuca gotowej decyzji i nie zamienia uznania administracyjnego w bezrefleksyjne zatwierdzanie rekomendacji. W epoce agentowej asystent przestaje być tylko doradcą: może przygotować kompletny projekt czynności, który trafia do warstwy kontrolnej.

Funkcja asystenta	Wartość	Zabezpieczenie
Wyszukiwanie prawa i procedur	szybszy dostęp do aktualnej podstawy	cytowanie wyłącznie z kanonicznego korpusu
Streszczanie akt	krótszy czas zapoznania się ze sprawą	lista pominiętych źródeł i poziom pewności
Kontrola kompletności	mniej wezwań i błędów formalnych	reguły jawne, możliwość ręcznej korekty
Projekt odpowiedzi	spójny język i standard obsługi	oznaczenie treści AI, akceptacja autora
Terminy i workflow	mniej przewlekłości i zagubionych spraw	log, eskalacja i właściciel procesu
Analiza wariantów	lepszą jakość argumentacji	obowiązek pokazania alternatyw i ryzyk

Produktywność musi być mierzona telemetrycznie. Deklarowana oszczędność czasu bywa wielokrotnie wyższa od efektu rejestrowanego w logach, a w części zadań AI spowalnia doświadczonych pracowników. Każde wdrożenie ma baseline, grupę porównawczą oraz pomiar jakości, nie tylko satysfakcji. Destylat potwierdza też, że zaufanie buduje się przez zaangażowanie personelu w testowanie modelu hybrydowego, a nie przez odgórne wdrożenie.

5.2. Warstwa kontrolna agentów (Organizational Control Layer)

Warstwa kontrolna jest najważniejszym nowym narzędziem administracji epoki agentowej. Przechwytuje każdą propozycję czynności agenta, zanim zostanie wykonana, porównuje ją z mandatem, uprawnieniami, regułami i kontekstem sprawy, a następnie kwalifikuje do jednej z czterech ścieżek. Każda kwalifikacja trafia do dziennika audytowego.

Ścieżka	Kiedy	Skutek
Zatwierdź (Approve)	czynność mieści się w mandacie, limitach i regułach	wykonanie z zapisem w dzienniku
Popraw (Revise)	czynność jest dopuszczalna co do celu, lecz wadliwa co do parametrów, podstawy lub zakresu	ustrukturyzowana naprawa propozycji i ponowna kwalifikacja
Zablokuj (Block)	czynność narusza zakaz, przekracza uprawnienia lub dotyczy obszaru niedelegowalnego	brak wykonania, informacja zwrotna dla agenta i właściciela
Eskaluj (Escalate)	niepewność, konflikt reguł, wysoki wpływ lub nowa klasa sytuacji	przekazanie do człowieka z pełnym kontekstem

Wynik badań jest jednoznaczny: warstwa kontrolna zredukowała niebezpieczne wykonania z 88% do zera, lecz samo blokowanie odzyskiwało tylko około 60% skuteczności zadań. Administracja, która traktuje kontrolę jako zakaz, płaci za bezpieczeństwo paraliżem; administracja, która wbudowuje naprawę, zachowuje zarówno bezpieczeństwo, jak i sprawność.

Governance agentów musi być fazowe: dowód koncepcji, eksperyment, produkcja — z człowiekiem jako finalnym arbitrem ryzyka w kluczowych punktach, a nie procesorem każdego wniosku. Warstwa kontrolna obejmuje również komunikację między agentami administracji: każda wiadomość A2A przenosi tożsamość nadawcy, mandat, właściciela kontekstu i atestowane zdolności, a nie samodeklaracje.

ZASADA WYKONAWCZA

Żaden agent administracji nie wykonuje czynności wobec obywatela, rejestru lub zasobu publicznego z pominięciem warstwy kontrolnej. Brak ścieżki naprawy jest wadą projektową, nie cechą bezpieczeństwa.

5.3. Elektroniczne zarządzanie dokumentacją, Narodowy Graf Wiedzy i RAG

EZD jest kręgosłupem, a nie archiwum po zakończeniu sprawy. System automatycznie rozpoznaje dokument, klasyfikuje go, wydobywa terminy i obowiązki, sprawdza kompletność, zapisuje proveniencję i zachowuje pełny ślad zmian. Dopiero na takiej warstwie można bezpiecznie budować RAG i agentów wiedzy.



Rysunek 8. Architektura Narodowego Grafu Wiedzy i RAG dla administracji.

Prawo i instrukcje są zmienne w czasie. Każdy fakt prawny ma dwie osie czasu: kiedy obowiązywał w świecie oraz kiedy system wiedział o zmianie. Bez bi-temporalności nie da się odtworzyć stanu prawnego na dzień wydania decyzji. Ten sam korpus kanoniczny, udostępniony publicznie przez API, jest źródłem dla agentów obywatelskich — dzięki temu obie strony sporu cytują ten sam, wersjonowany stan prawa.

Moduł	Wymóg minimalny
Dokument AI	OCR/HTR, analiza układu, klasyfikacja, ekstrakcja i walidacja.
EZD	retencja, metadane, wersjonowanie, legal hold, ścieżka audytu.
Korpus kanoniczny	autorytatywne źródła, daty obowiązywania, historia zmian, publiczne API.
RAG/GraphRAG	wyszukiwanie hybrydowe, cytowanie, wynik ugruntowania (groundedness).
Pamięć instytucjonalna	agent skanujący archiwa i raporty resortowe, wydobywający reużywalne wzorce do centralnego repozytorium.
Odpowiedź	źródło, data, wersja dokumentu, poziom pewności i ograniczenia.
Bezpieczeństwo	ochrona przed zatruciem wiedzy, prompt injection i wyciekiem kontekstu.

Pamięć instytucjonalna odpowiada na tzw. paradoks Salomona: instytucja nie rozpoznaje wiedzy, którą już posiada. Generatywna AI działająca jako transduktor funkcji S4 może wydobywać z rozproszonych archiwów rozwiązania, które zostały już raz wypracowane, i udostępniać je innym jednostkom.

5.4. AI w legislacji, podatkach, zamówieniach i audycie

Obszar	Zastosowanie	Warunek bezpieczeństwa
Legislacja	kontrola spójności, konsolidacja, analiza skutków, mapowanie konsultacji, rules as code	kanoniczny korpus, jawne źródła, ludzka odpowiedzialność autora
Podatki	walidacja deklaracji, prognozowanie, wykrywanie anomalii	raport trafności i błędów fałszywie dodatnich, zakaz ukrytego profilowania
Zamówienia	asystent opisu przedmiotu zamówienia, analiza ofert, konflikt interesów, zmowa przetargowa	prawo audytu, klauzula wyjścia, ujawnienie podwykonawców i danych
Audyt	ciągły monitoring, process mining, dobór próby, wykrywanie nadużyć	niezależność audytora i odrębność modeli od systemów wykonawczych
Finanse publiczne	prognoza dochodów i wydatków, monitoring ryzyk, kontrola umów	miernik nie zastępuje celu publicznego ani praw obywatela

Zamówienia na AI wymagają reżimu zgłaszania wyzwalanego kategorią systemu, nie wartością umowy. Narzędzie darmowe, darowizna lub moduł dodany do istniejącej platformy może tworzyć większe ryzyko niż formalny przetarg. Każdy dostawca zapewnia audyt, wersjonowanie, logi, przenoszalność, migrację i możliwość utrzymania przez podmiot trzeci. W legislacji automatyczna wizualizacja debat parlamentarnych i redagowanie kart usług prostym językiem z walidacją ludzką są narzędziami inkluzji: dostępność oznacza zrozumiałość, nie tylko formalną publikację.

5.5. Decyzje wielokryterialne: AHP jako tłumacz ocen modeli

Modele językowe używane jako oceniające mają skłonność do tendencji centralnej: ich surowe oceny skupiają się wokół środka skali. Przepuszczenie różnic między ocenami (mierzonych dystansem Jensena–Shannona) przez macierz porównań parami AHP daje wagi interpretowalne i odporne na skrzywienie skali. Destylat opisuje kompletny wzorzec RAD (Retrieval-Augmented Decision-making).

Krok	Czynność	Funkcja kontrolna
1	model wydobywa kryteria z korpusu dokumentów programowych i prawnych	kryteria mają źródło, nie są wymyślone
2	hierarchizacja kryteriów metodą ISM	jawna struktura zależności
3	wagi z wieloagentowego „głosu” symulowanych ekspertów i AHP	spójność mierzona wskaźnikiem CR
4	ocena wariantów z kontrolą spójności w czasie rzeczywistym	indeks Koczkodaja wskazuje, które porównanie jest sprzeczne
5	raport uzasadnień z pełną ścieżką audytu	decyzja odtwarzalna przez osobę trzecią

Wskaźnik spójności przestaje być formalnością po fakcie: staje się detektorem jakości danych, który natychmiast prosi o rewizję konkretnego porównania. W grupowym AHP ten sam aparat wykrywa manipulację: podejrzani są zarówno uczestnicy odstający rankingiem, jak i ci, których oceny są nienaturalnie spójne lub niespójne. Ich wagi są automatycznie obniżane bez jawnego wykluczania. To bezpośrednia ochrona konsultacji obywatelskich i budżetów partycypacyjnych przed astroturfingiem.

ZASADA JEDNEGO FRAMEWORKU

Doświadczenie ośmiu gmin holenderskich pokazuje, że zautomatyzowane podejmowanie decyzji bez odgórnego standardu kończy się chaosem wdrożeń ad hoc. Narzędzia wielokryterialne administracji wymagają jednego centralnego frameworku governance, a nie projektu po projekcie.

5.6. Systemy wsparcia decyzji z zachowaniem autonomii decydenta

Systemy wsparcia decyzji w sądownictwie i administracji mogą podnosić jakość rozstrzygnięć bez odbierania decydentowi autonomii. System JUSTICE realizuje proces „wyszukaj — wstępnie oceń — napisz”; pominięcie etapu wstępnej oceny obniża trafność, co pokazuje, że struktura procesu jest ważniejsza od mocy modelu. System ARGOS buduje model per decydent, a nie model globalny: zasada niezawisłości wymaga personalizacji wsparcia, a nie centralizacji rozstrzygnięć.

W zarządzaniu kryzysowym wsparcie decyzji wymaga formalnie typowanych interfejsów narzędzi i geoprzestrzennych priorów analogowych, które utrzymują konserwatywną linię bazową przy ubogich danych. Trzy grupy interesariuszy krytykują przy tym różne etapy cyklu życia systemu: obywatel pyta o cel, ekspert techniczny o metodę, funkcjonariusz o dopasowanie do praktyki. Interfejs ewaluacji musi umożliwiać różne pytania dla różnych ról.

5.7. AIOps i zarządzanie infrastrukturą państwa

AIOps wykorzystuje telemetrię, logi, modele anomalii i agentów operacyjnych do utrzymywania systemów państwa. Największą wartością nie jest pełna automatyzacja, lecz skrócenie czasu wykrycia, diagnozy i bezpiecznego odtworzenia usługi. W infrastrukturze krytycznej każda automatyczna remediacja ma limit, tryb cienia, rollback i wyłącznik awaryjny.

Funkcja AIOps	Przykład	Poziom autonomii
Obserwowalność	korelacja logów, zdarzeń i metryk	A0–A1
Triage incydentów	deduplikacja alertów i priorytetyzacja	A1–A2
Analiza przyczyny	propozycja przyczyny źródłowej i zależności	A1
Remediacja	restart, izolacja, zmiana konfiguracji	A2 tylko z rollbackiem
Pojemność i koszt	prognoza obciążenia, autoskalowanie, FinOps	A2–A3 w limitach
Cyberoperacje	wykrywanie anomalii tożsamości i zachowania	A1–A2; krytyczne akcje zatwierdza człowiek

AIOps nie może tworzyć drugiej, niekontrolowanej warstwy administracji. Agenci operacyjni potrzebują własnej tożsamości maszynowej, krótkotrwałych poświadczeń, listy dozwolonych narzędzi, budżetu działań i pełnego śladu. W epoce agentów obywatelskich AIOps obejmuje także wykrywanie zalewu wniosków i odróżnianie go od rzeczywistego wzrostu potrzeb.

ODPORNOŚĆ

System krytyczny jest gotowy dopiero wtedy, gdy potrafi działać w trybie degradacji, odtworzyć się z kopii zapasowej i przejść na procedurę ręczną po odcięciu usług zewnętrznych.

Podstawa: destylat NAI, soczewki I i II; wydanie pierwsze „Narzędzia AI”, rozdz. 4, 7–9; TS-CC Controlling cybernetyczny; TS-NSB Narodowa Strategia Bezpieczeństwa; publikacja 46 „Analityczny proces wyboru technologii militarnych”.

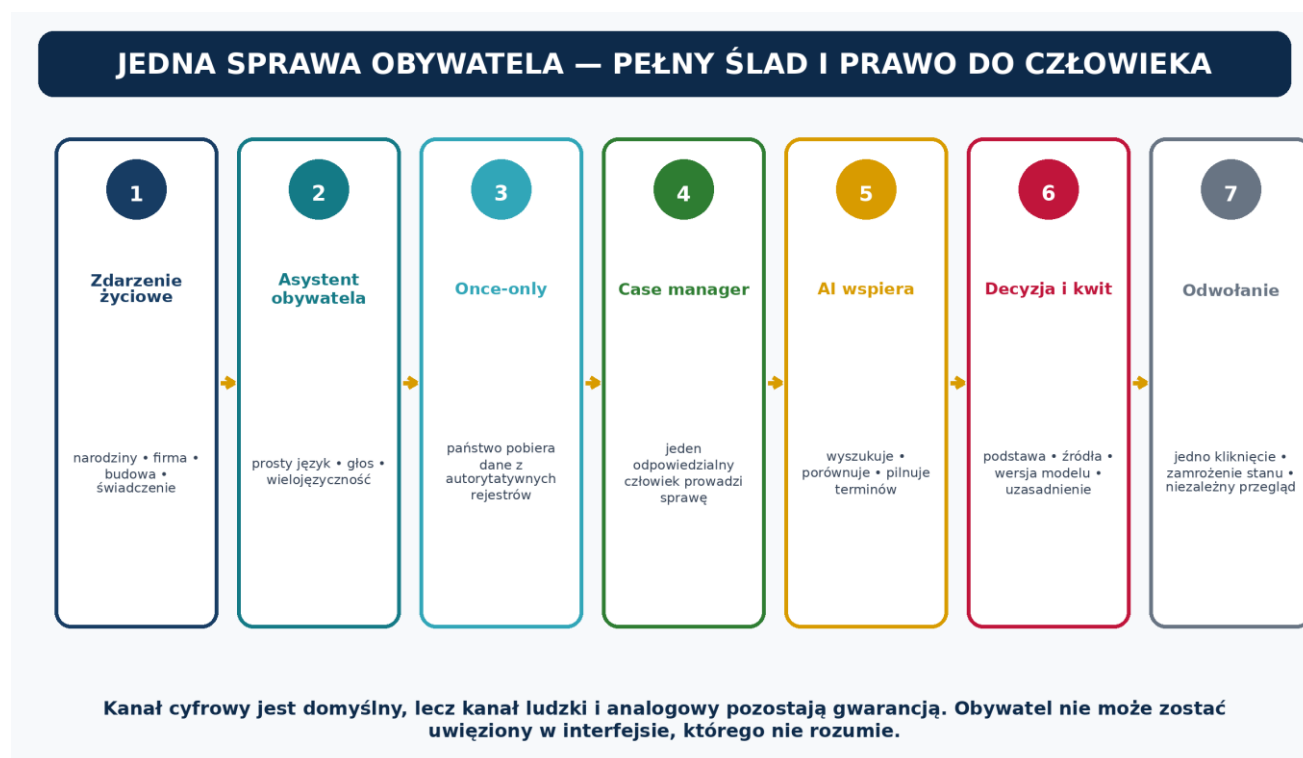
6. Narzędzia AI obywatela

Podmiotowość realna, nie deklaratywna

Rozdział ten jest nowy i priorytetowy. Odpowiada na pytanie, jakie narzędzia AI dać obywatelowi, aby jego podmiotowość wobec państwa była realna. Klaster CITIZENTOOLS jest w destylacie klastrem priorytetowym, a jego wnioski są zarazem najbardziej optymistyczne i najbardziej ostrzegawcze: narzędzia obywatelskie działają, lecz tylko w sprzężeniu z odpowiednim zobowiązaniem państwa.

6.1. Jedna brama usług i zdarzenia życiowe

Obywatel nie powinien znać mapy urzędów. Opisuje potrzebę, a państwo rozpoznaje zdarzenie życiowe, wskazuje właściciela sprawy, pobiera dane, które już posiada, proponuje ścieżkę i pokazuje status. Brama jest pojedynczym punktem wejścia, lecz za nią działa federacja wyspecjalizowanych agentów.



Rysunek 9. Ścieżka sprawy obywatela z asystentem, case managerem, Kwitem Decyzji i prawem do człowieka.

Standard usługi	Treść operacyjna
One-stop government	jedna brama cyfrowa i fizyczna niezależnie od liczby instytucji w tle
Once-only	organ nie żąda danych, które państwo już posiada w autorytatywnym rejestrze
Zdarzenia życiowe	usługa projektowana wokół sytuacji człowieka, nie procedury urzędu
Interfejs dla agenta	każda usługa ma maszynowo czytelny opis, schemat wniosku, statusu i uzasadnienia
Prosty język	odpowiedź zrozumiała bez prawniczego pośrednika, testowana zadaniowo
Human handoff	natychmiastowe przejście do człowieka bez utraty historii rozmowy
Kanał analogowy	technologia nie wyklucza osoby, która nie korzysta z internetu
Status i SLA	obywatel widzi termin, odpowiedzialną osobę, brakujące dane i drogę eskalacji

DLA PRZEDSIĘBIORCY

Jedno państwo oznacza jeden standard pozwolenia, podatku, kontroli i zamówienia. Agent przedsiębiorcy, który komunikuje się z maszynowo czytelnym interfejsem państwa, skraca permitting bardziej niż kolejna nowelizacja terminów.

6.2. Agent obywatelski i problem zalewu wniosków

Agent obywatelski redukuje trzy typy obciążeń administracyjnych opisane przez Moynihana, Herd i Harveya: koszty poznawcze (zrozumienie, co i gdzie załatwić), koszty zgodności (formularze, dokumenty, terminy) oraz koszty psychologiczne (stres, stygmatyzacja, poczucie bezradności). Dla osób starszych, z niepełnosprawnościami, o niskich kompetencjach cyfrowych i dla migrantów wewnętrznych jest to zmiana o znaczeniu porównywalnym z wprowadzeniem bezpłatnej pomocy prawnej.

Ten sam mechanizm generuje jednak zjawisko agentic flooding: 84 udokumentowane przypadki przeciążenia usług w 11 krajach. Jeżeli państwo nie przygotuje pojemności i interfejsów, odpowie opłatami, limitami i barierami, które uderzą najmocniej w najślabszych — czyli w tych, dla których agent miał być wyrównaniem szans.

Warunek wdrożenia agenta obywatelskiego	Mechanizm
Reforma pojemności usług	prognoza obciążenia, elastyczne zasoby, automatyczne rozpatrywanie spraw prostych po stronie państwa
Maszynowo czytelne zobowiązania	DSAR endpoint, schemat wniosku i odpowiedzi, termin, format uzasadnienia
Tożsamość i delegacja	agent działa na podstawie jawnej, odwołalnej delegacji od unikalnej osoby
Deduplikacja i łączenie	wiele wniosków w tej samej sprawie łączonych w jedną sprawę bez utraty praw
Zakaz barier dyskryminujących	limity techniczne nie mogą ograniczać praw osób działających przez agenta

Granica: pełnomocnik, nie wyborca

Idea „jedna osoba — jeden bot” jako pełnomocnika politycznego jest atrakcyjna, lecz niebezpieczna bez trzech zabezpieczeń: twardego podziału na decyzje konstytucyjne i bieżące, jawności źródeł profilu preferencji oraz rozwiązania problemu, kto strzeże infrastruktury roju agentów. Dlatego w TechnoRepublice zakres agenta obywatelskiego jest doradczy i wykonawczy — agent może przygotować, złożyć, przypomnieć, porównać i wyjaśnić — lecz nigdy nie głośnie w imieniu obywatela na poziomie ustrojowym, w referendum ani w wyborach.

6.3. Portfel danych i samosuverenna tożsamość

Stos „obywatel jako właściciel własnych danych” składa się z czterech elementów: osobistego skarbcza danych (Personal Data Vault), samosuverennej tożsamości (SSI), weryfikowalnych poświadczeń (Verifiable Credentials) i żądania dostępu do danych jednym kliknięciem (one-click DSAR). Stos ten obsługuje nie tylko relację obywatel–korporacja, ale przede wszystkim relację obywatel–państwo.

Funkcja	Co zyskuje obywatel	Zobowiązanie państwa
Log dostępu	widzi, który organ i w jakim celu pobrał jego dane	każdy odczyt z rejestru jest logowany i udostępniany właścicielowi danych
Selektywne ujawnianie	przekazuje tylko atrybut potrzebny do sprawy	usługi weryfikują atrybut, nie cały dokument
One-click DSAR	pobiera kopię swoich danych i profili w formacie maszynowym	ustandaryzowany endpoint i termin odpowiedzi
Korekta danych	zgłasza błąd z dowodem w portfelu	procedura korekty z terminem i propagacją do rejestrów zależnych
Delegacja	udziela agentowi lub osobie zaufanej zakresowego dostępu	weryfikacja delegacji i jej zakresu przy każdej czynności

6.4. Samoobrona prawna i deterministyczny walidator

AI jako narzędzie samoobrony prawnej działa formalnie, lecz zawodzi materialnie. Badania pokazują więcej złożonych pism, ale wyższy wskaźnik odrzuceń skarg oznaczonych jako przygotowane przez AI. Przyczyny są dwie: halucynowane lub nieadekwatne cytowania oraz trening na ogólnym korpusie prawa zamiast na sprawach, z którymi rzeczywiście mierzą się osoby o niskim statusie społeczno-ekonomicznym.

Wymóg publicznego narzędzia samoobrony	Uzasadnienie
deterministyczny walidator cytowań	każdy przepis i orzeczenie sprawdzane w kanonicznym, bi-temporalnym korpusie
korpus spraw typowych dla osób najbliższych	świadczenia, eksmisje, zadłużenie, kontrole, sprawy pracownicze
szacowanie szans i kosztów	obywatel wie, czy pismo ma sens, zanim je złoży
przekierowanie do pomocy ludzkiej	nieodpłatna pomoc prawna, rzecznik, organizacja społeczna
oznaczenie udziału AI bez stygmatyzacji	organ nie może traktować pisma gorzej z powodu użycia narzędzia

6.5. Obywatelski audyt algorytmów

Obywatelski audyt algorytmów działa w dwóch komplementarnych trybach. Tryb spontaniczny — amplifikacja i eskalacja przypadków krzywdy w mediach i sieciach — jest realnym mechanizmem nacisku, co pokazały afera zasiłkowa w Holandii i orzeczenie w sprawie SyRI. Tryb strukturalny wymaga zestawu narzędzi, a jego pierwszym elementem jest rozpoznanie, że dany system w ogóle jest systemem algorytmicznym. Nawet eksperci nie rozpoznają modelu uczenia maszynowego jako „systemu algorytmicznego” bez narzędzia. Rozpoznanie jest więc warunkiem wstępnym rejestru, a nie jego następstwem.

Po stronie prawa punkt wyjścia jest w Polsce solidny: algorytm stosowany przez administrację jest informacją publiczną, a Trybunał Konstytucyjny w wyroku K 53/16 zakwestionował profilowanie bezrobotnych oparte na niejawnych zasadach. Destylat wskazuje jednak, że prawo do wyjaśnienia logiki — a nie do surowego kodu — jest właściwym narzędziem dostępu, i że oba obowiązki trzeba prawnie rozdzielić.

OBYWATELSKI TOOLKIT AUDYTU

Kwestionariusz rozpoznania systemu algorytmicznego; czytnik rejestru algorytmów przez API; generator wniosku o wyjaśnienie logiki decyzji; zestaw testów porównawczych dla spraw podobnych; szablon skargi do KRIBSI i PUODO; publiczny rejestr zgłoszonych przypadków.

6.6. Prosty język i dostępność jako infrastruktura

Prosty język jest infrastrukturą, a nie stylem. Standardowe metryki czytelności zawodzą dla tekstu prawnego, ponieważ jego trudność ma charakter strukturalny, a nie leksykalny. Właściwą metodą jest ewaluacja zadaniowa: czy obywatel po przeczytaniu poprawnie odpowiada na pytanie „czy mam prawo do X i co mam zrobić”. Dostępność cyfrowa jest przy tym luką kompetencyjną urzędników, a nie technologii: najbardziej pożądaną funkcją systemu zarządzania treścią jest automatyczne dopasowanie do WCAG — zadanie dla wbudowanego AI, a nie kolejnego szkolenia.

6.7. Partycypacja, synteza konsensusu i ochrona konsultacji

AI wzmacnia partycypację w dwóch sprawdzonych trybach: dostępności informacji (wyjaśnienie projektu ustawy, budżetu, planu miejscowego w języku obywatela) oraz syntezy konsensusu (odnajdywanie stanowisk, które akceptują grupy o różnych poglądach). Bez publicznej, a nie rynkowej infrastruktury narzędzia te odtworzą nierówności cyfrowe — zjawisko określane jako *agentic inequality*: lepsi agenci dla bogatszych obywateli, organizacji i lobbystów.

Narzędzie partycypacji	Zabezpieczenie
konsultacje z syntezą stanowisk	dowód unikalności uczestnika; publikacja metody syntezy
budżet partycypacyjny wspierany AI	grupowy AHP z wykrywaniem manipulacji i spójności
petycje i inicjatywy obywatelskie	personhood credential; jawny licznik unikalnych podpisów
wyjaśnianie projektów aktów prawnych	wersja zadaniowo testowana prostym językiem, źródło w korpusie kanonicznym
wizualizacja debat i głosowań	pełne dane otwarte, bez selekcji i komentarza

6.8. Sygnaliści i kanał zwrotny z obowiązkiem reakcji

W przypadku sygnalistów anonimowość jest z definicji niewystarczająca. Zjawisko określane jako epistemiczna niedeanonimizowalność oznacza, że sama treść zgłoszenia często wystarcza do identyfikacji autora. Ochrona wymaga trójkąta: technologii (bezpieczny kanał, minimalizacja metadanych, AI redagujące treść bez utraty istoty), prawa (zakaz odwetu, odwrócony ciężar dowodu) i kultury organizacyjnej. Każdy kanał zwrotny — skarga, sygnał, zgłoszenie usterki — działa tylko sprzężony z instytucją, która ma obowiązek i zdolność zareagować.

6.9. Asystent dla osób wykluczonych cyfrowo

Asystent dla grup wykluczonych cyfrowo pełni inną funkcję niż chatbot transakcyjny. Destylat opisuje go metaforą przewodnika po lesie z jadowitymi roślinami: celem nie jest przepustowość, lecz bezpieczna autonomia. Asystent działa głosem, prostym językiem, w tempie użytkownika, przypomina, ostrzega przed oszustwem, pozwala w każdej chwili przejść do człowieka i współpracuje z osobą zaufaną — członkiem rodziny, mentorem z Programu 2G lub pracownikiem terenowego centrum obsługi.

ZASADA INKLUZJI

Narzędzia AI obywatela są dobrem publicznym. Ich jakość nie może zależeć od dochodu, wieku, miejsca zamieszkania ani kompetencji cyfrowych użytkownika.

Podstawa: destylat NAI, soczewki III i V; Moynihan, Herd, Harvey (2015); Adler i in. (2024) — personhood credentials; TK K 53/16; Rechtbank Den Haag, SyRI (2020); eIDAS 2.0.

7. Prawa obywatela, algorytmiczny due process i Public Constitutional AI

AI ma wzmacniać proceduralną sprawiedliwość, nie ją obchodzić

7.1. Prawo do dobrej administracji w epoce agentów

Prawo do dobrej administracji obejmuje legalność, bezstronność, proporcjonalność, terminowość, czynny udział strony, dostęp do akt, obowiązek uzasadnienia i skuteczny środek odwoławczy. Użycie AI nie zmniejsza tych obowiązków. Przeciwnie: wymaga dodatkowej przejrzystości, ponieważ obywatel nie widzi sposobu działania modelu ani przepływu danych. Epoka agentów dodaje nowe prawa, które wynikają z zasady symetrii.

Prawo obywatela	Standard TechnoRepubliki
Prawo do informacji	jasne oznaczenie użycia AI i wskazanie odpowiedzialnego organu (zgodnie z art. 50 AI Act).
Prawo do bycia wysłuchanym	możliwość przedstawienia kontekstu, którego nie ma w danych.
Prawo do uzasadnienia	podstawa prawna, fakty, źródła, rola AI i rozumowanie człowieka — w Kwicie Decyzji.
Prawo do korekty danych	szybka procedura poprawy błędnego rejestru lub flagi, z propagacją.
Prawo do przeglądu przez człowieka	realny, niezależny i kompetentny przegląd.
Prawo do odwołania	skuteczny środek prawny, zamrożenie stanu i zachowanie logów.
Prawo do własnego agenta	państwo nie może odmówić obsługi sprawy prowadzonej przez agenta działającego na podstawie jawnej delegacji.
Prawo do maszynowego interfejsu	usługa publiczna ma maszynowo czytelny opis, status i format odpowiedzi.
Prawo do logu dostępu	obywatel widzi, kto i w jakim celu przetwarzał jego dane.
Prawo do kanału niecyfrowego	obsługa przez człowieka dla osób wykluczonych technologicznie.
Prawo do niedominacji	zakaz ukrytej manipulacji, social scoringu i niezaskarżalnej władzy algorytmicznej.

ALGORYTMICZNY DUE PROCESS

Wyjaśnienie + możliwość zakwestionowania + przegląd przez człowieka + niezależna kontrola + skuteczna naprawa szkody. Każdy z elementów ma w TechnoRepublice swoje narzędzie po stronie obywatela i swoje zobowiązanie po stronie administracji.

7.2. Wyjaśnienia dla trzech odbiorców

Wyjaśnialność musi być zróżnicowana według trzech grup odbiorców, ponieważ każda zadaje fundamentalnie inne pytanie. Jedna metoda wyjaśnień nie odpowiada na wszystkie.

Odbiorca	Pytanie	Forma wyjaśnienia
Obywatel	dlaczego ja, co mogę zmienić, jak się odwołać	Kwit Decyzji z czynnikami wpływu i kontrfaktykiem
Regulator i audytor	czy system jest zgodny z prawem i nie dyskryminuje	karta systemu, dokumentacja techniczna, testy grupowe, logi
Zespół techniczny	dlaczego model zachował się w ten sposób	atrybucja cech, analiza błędów, monitoring dryfu

7.3. Public Constitutional AI

Zasady, którymi kieruje się AI w państwie, nie mogą być wyłącznie dokumentem przyjętym przez dostawcę lub resort. Legitymizacja wymaga procesu, który destylat określa jako Public Constitutional AI. Składa się on z trzech odrębnych narzędzi — a nie z jednego dokumentu zasad.

Narzędzie	Funkcja	Instytucja w TechnoRepublice
AI Constitution-Making	obywatele współtworzą reguły nadrzędne dla systemów AI państwa	konsultacje z syntezą konsensusu, Senat 100, referendum w sprawach ustrojowych
AI Courts	mechanizm kontestacji konkretnej reguły i jej zastosowania	odwołanie, skarga do KRiBSI, kontrola sądowa, Sąd Najwyższy
AI Compliance	widoczne wymuszanie zgodności systemów z regułami	rejestr algorytmów, audyt niezależny, publiczny pulpit KPI/KRI

Public Constitutional AI jest praktycznym wykonaniem funkcji S5 modelu systemu żywego: wartości nadrzędne nie są zapisane raz, lecz ustanawiane, kwestionowane i egzekwowane w pętli, w której uczestniczy suweren.

7.4. Człowiek w pętli jako obowiązek prawny

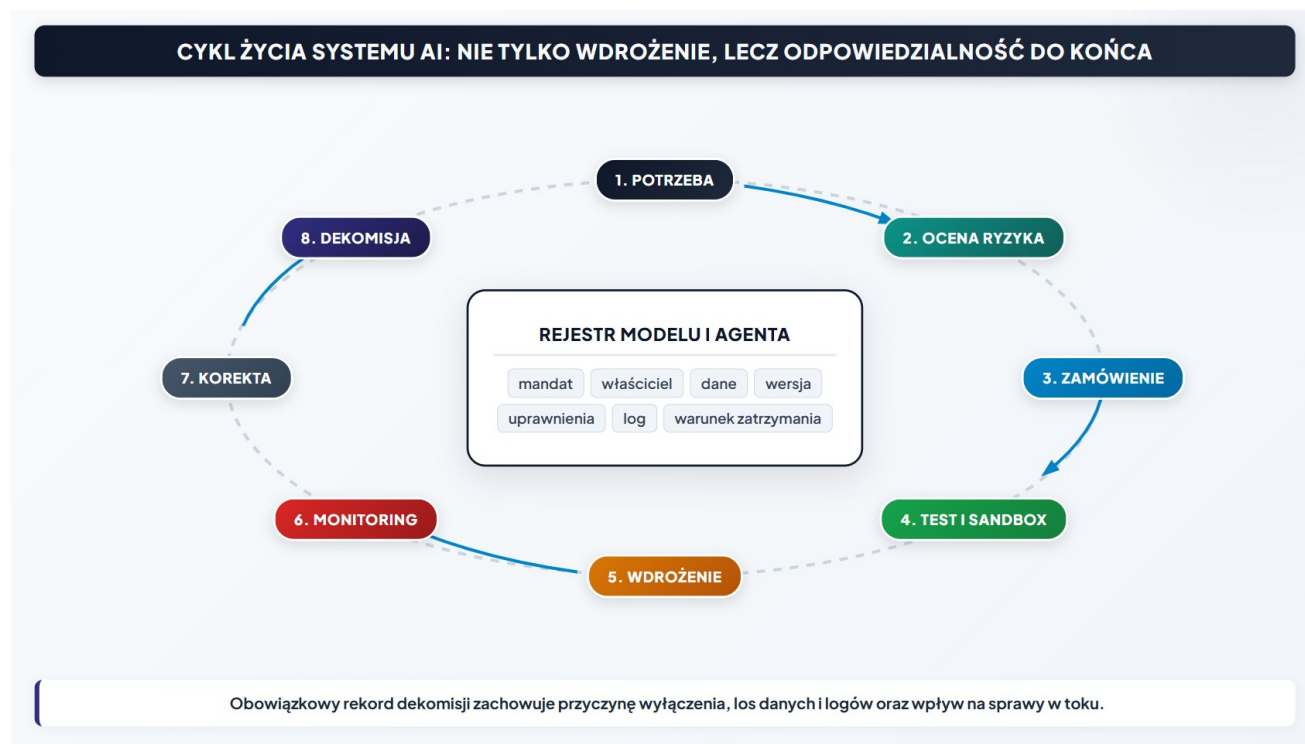
„Człowiek w pętli” nie może być ogólnym sloganem. Ustawa i karta systemu określają kompetencje, czas, zakres informacji, prawo do zatrzymania, ścieżkę eskalacji i odpowiedzialność osoby sprawującej nadzór. Obywatel zna tożsamość organu, niekoniecznie nazwisko każdego operatora, ale zawsze istnieje człowiek i instytucja zdolna odpowiedzieć oraz naprawić błąd. Orzeczenie TSUE w sprawie SCHUFA (C-634/21) potwierdziło, że już samo generowanie wyniku, od którego w decydującym stopniu zależy decyzja, może być zautomatyzowanym podejmowaniem decyzji w rozumieniu art. 22 RODO.

Podstawa: destylat NAI IV.6–8, V.2, V.10; art. 22 RODO; art. 41 Karty Praw Podstawowych UE; art. 50 i 86 AI Act; TSUE C-634/21; projekt Konstytucji TechnoRepubliki Polskiej.

8. Zarządzanie modelami, agentami i danymi

Rejestr, wersja, mandat, log i dekomisja

System AI nie kończy się w dniu odbioru. Model może się zmienić, dane mogą się zestarzeć, kontekst może zostać zatruty, a dostawca może uruchomić nową funkcję. Administracja potrzebuje cyklu życia obejmującego zatwierdzenie, test, wdrożenie, obserwację, korektę, zawieszenie i dekomisję.



Rysunek 10. Cykl życia systemu AI z rejestrem, monitoringiem i obowiązkową dekomisją.

Element karty systemu	Treść
Identyfikator i właściciel	jednoznaczna nazwa, jednostka, osoba odpowiedzialna, dostawca.
Mandat i cel	podstawa prawna, dozwolone i zakazane zastosowania.
Para sprzężona	narzędzie obywatela umożliwiające zrozumienie, użycie lub zakwestionowanie systemu.
Dane	źródła, podstawa przetwarzania, jakość, retencja, lineage.
Model i wersja	wagi lub API, data wdrożenia, benchmark, ograniczenia, historia zmian.
Autonomia i narzędzia	poziom A0–A5, lista dozwolonych narzędzi, limity kosztu, uprawnienia.
Komunikacja A2A	dozwoleni partnerzy, propagacja tożsamości i mandatu, atestacja zdolności.
Nadzór	role, limit obciążenia, bramki, ścieżka eskalacji, wyłącznik awaryjny.
Monitoring	KPI, KRI, dryf, incydenty, harmonogram przeglądów, test regresji wartości.
Dekomisja	warunek zatrzymania, migracja, los danych i logów, sprawy w toku.

8.1. Rejestr algorytmów przeciw teatrowi przejrzystości

Destylat pokazuje, że rejestry algorytmów systematycznie stają się teatrem przejrzystości z powodu trzech luk: zakresu (wybiórcza rejestracja), głębokości (wolny tekst bez struktury) i wymuszalności (dobrowolność). Działają wtedy jako wewnętrzny instrument zarządzania, lecz nie jako zewnętrzna rozliczalność, bo nie mają krytycznego odbiorcy. Rejestr TechnoRepubliki zamyka wszystkie trzy luki.

Luka	Objaw	Rozwiązanie
Zakres	rejestruje się systemy „wizerunkowe”, pomija ukryte moduły i narzędzia darmowe	obowiązek wyzwalany kategorią systemu; brak wpisu = zakaz użycia
Głębokość	opisy prozą bez porównywalnych pól	schemat danych, słowniki kontrolowane, wersjonowanie, API
Wymuszalność	brak sankcji za nieaktualny wpis	KRI „brak aktualizacji po zmianie wersji”, sankcja i odpowiedzialny organ
Odbiorca	nikt nie analizuje rejestru	obywatelski toolkit audytu, dostęp dla rzecznika, UODO, sądów i organizacji watchdog

Podstawa: destylat NAI IV.7, V.4; wydanie pierwsze „Narzędzia AI”, rozdz. 10; ISO/IEC 42001:2023; NIST AI RMF 1.0.

9. Bezpieczeństwo, odporność i suwerenność

Najpierw ogranicz uprawnienia, potem zwiększ inteligencję

9.1. Nowe klasy ryzyka agentów



Rysunek 11. Warstwowy model bezpieczeństwa i suwerenności administracji AI.

Agenci AI łączą model, pamięć, narzędzia i uprawnienia. To tworzy nowe klasy ryzyka: prompt injection, przejęcie narzędzi, zatrucie pamięci, eksfiltrację danych, podszywanie się, pętle delegacji, awarie kaskadowe i ataki na budżet obliczeniowy. Zabezpieczenia muszą działać w czasie wykonania, nie tylko w dokumentacji przed wdrożeniem.

Ryzyko	Minimalna odpowiedź
Prompt/tool injection	separacja instrukcji i danych, lista dozwolonych narzędzi, proxy narzędzi, walidacja wyniku.
Nadmierne uprawnienia	least privilege, krótkie tokeny, dostęp związany z celem, segregacja obowiązków.
Luki protokołów A2A	powiązanie kontekstu z właścicielem, propagacja tożsamości w łańcuchu delegacji, atestacja zdolności.
Wyciek danych	minimalizacja, pseudonimizacja, izolacja najemców, poufne przetwarzanie.
Halucynacja prawna	RAG z korpusu kanonicznego, weryfikator cytowań, odmowa odpowiedzi przy braku źródła.
Zatrucie wiedzy	proweniencja, podpisy, kwarantanna źródeł, wersjonowanie i detekcja anomalii.
Awaria kaskadowa	limity delegacji, timeout, wyłącznik obwodu, tryb bezpieczny i ręczny.
Zalew wniosków	deduplikacja, tożsamość delegacji, prognoza obciążenia, łączenie spraw.
Deepfake i podszywanie	poświadczenia pochodzenia treści, znakowanie, potwierdzenie tożsamości kanałem niezależnym.
Shadow AI	polityka dopuszczalnego użycia, zatwierdzone narzędzia, DLP i szkolenia.

Architektura bezpieczeństwa sektora publicznego musi przy tym uwzględniać realny profil zagrożeń. Według danych destylatu 63% incydentów w sektorze publicznym to hacktywizm (głównie ataki DDoS), a zaawansowane kampanie państwowe stanowią około 2,5%. Masowy, nieskomplikowany atak jest codziennością i musi być obsługiwany automatycznie; zasoby ludzkie należy koncentrować na rzadkich atakach zaawansowanych.

Dyrektywy CER i NIS2 są prawnie uzupełniające, a nie zamienne: pierwsza dotyczy odporności fizyczno-operacyjnej podmiotów krytycznych, druga cyberbezpieczeństwa. System AI w infrastrukturze krytycznej musi być klasyfikowany pod obie jednocześnie, w sposób skoordynowany.

ZASADA ZERO TRUST

Każdy agent, narzędzie, źródło i żądanie są uwierzytelniane oraz ograniczane. Zaufanie nie wynika z tego, że system działa „wewnątrz urzędu” — ani z tego, że agent deklaruje swoje zdolności.

9.2. Suwerenność jako zarządzana współzależność

Pełna suwerenność AI jest strukturalnie niewykonalna: żadne średnie państwo nie kontroluje całego łańcucha od litografii po modele graniczne. Realną strategią jest zarządzana współzależność: zmapować zależności, priorytetyzować tam, gdzie kraj ma realną przewagę — w przypadku Polski są to dane, jakość governance i talent, a nie produkcja sprzętu — oraz dywersyfikować dostawców. Suwerenna chmura ma trzy niezależne wymiary: operacyjny, danych i oprogramowania. Deklaracja suwerenności bez rozbicia na te trzy wymiary jest marketingiem.

Warstwa	Cel suwerenności	Minimalny standard
Dane	kontrola lokalizacji, celu i retencji	katalog, lineage, szyfrowanie, klucze kontrolowane przez państwo
Model	kontrola wersji, ograniczeń i zmian	otwarte wagi lub kontraktowe zamrożenie wersji, benchmark, notyfikacja zmian
Prompt i wiedza	kontrola instrukcji i korpusu	repozytorium promptów, wersjonowanie, audyt źródeł
Chmura	ciągłość i przenoszalność	multi-sourcing, standardowe formaty, brak karnego kosztu wyjścia
Compute	zdolność krytycznych obliczeń	krajowa lub europejska rezerwa, tryb ograniczony i priorytety
Tożsamość	kontrola korzenia zaufania	krajowe i europejskie usługi zaufania, portfel EUDI
Dostawca	brak przechwycenia technologicznego	prawo audytu, know-how do utrzymania, plan wyjścia
Ekosystem	konkurencja i kompetencje	kilku niezależnych integratorów, wsparcie MŚP i otwartych modeli

REGUŁA ZAKUPU

Ogólny poziom suwerenności jest równy najniższej warstwie. Wysoka punktacja w jednej warstwie nie kompensuje braku prawa wyjścia, kontroli kluczy lub możliwości audytu.

Podstawa: destylat NAI IV.4–5, IV.9–10, I.4; OWASP Top 10 for LLM Applications 2025; dyrektywy (UE) 2022/2555 i 2022/2557; TS-NSB Narodowa Strategia Bezpieczeństwa TechnoRepubliki Polskiej; publikacja 45 „Infrastruktura odporności”.

10. Gotowość organizacyjna, kadry i Program 2G

Nie kupuj agentów do organizacji, która nie kontroluje procesów i danych

10.1. Poziomy gotowości

Gotowość obejmuje strategię, procesy, dane, technologię, kadry, prawo, zamówienia, bezpieczeństwo, kulturę uczenia się i właścicielstwo rezultatów. Wdrożenie w organizacji niedojrzalej cyfrowo wzmacnia istniejący chaos oraz zależność od dostawcy. Destylat przypomina, że AI w administracji nie może kopiować sektora prywatnego: inny cel — dobro publiczne — wymaga reformy organizacyjnej równoległej z wdrożeniem.

Poziom	Charakterystyka	Dopuszczalny zakres
0. Ad hoc	shadow AI, brak polityki, brak inwentarza	zakaz danych służbowych w narzędziach publicznych; edukacja i audyt
1. Świadomość	świadomość ryzyka, pierwsze zasady	sandbox, szkolenia, brak skutków prawnych
2. Pilotaż	właściciel, baseline, dane i procedura testowa	odwracalne zastosowania w środowisku kontrolowanym
3. Zarządzany	rejestr, ocena ryzyka, monitoring, audyt, warstwa kontrolna	wdrożenia produkcyjne A0–A2 z parą obywatelską
4. Skalowany	wspólna platforma, standardy, centra kompetencji, interfejsy dla agentów	procesy end-to-end i agentowe A2–A3
5. Adaptacyjny	ciągła ewaluacja, dekomisja, uczenie instytucjonalne	kontrolowana adaptacja A4 w wybranych domenach

Globalny indeks dojrzałości GovTech Banku Światowego obejmuje 48 wskaźników, lecz analiza z użyciem AI wyłoniła trzy krytyczne dla przeskoczenia klasy dojrzałości: usługi zorientowane na obywatela, jednolitą strategię narodową i przejrzystość. Te trzy należy ważyć wyżej niż pozostałe. Ciągła, automatyczna ocena transformacji — równoległe na danych ankietowych pracowników i treści stron internetowych urzędów — wychwytuje rozjazd między deklaracją a stanem faktycznym lepiej niż coroczny ręczny audyt.

10.2. Kadry i nowe role

Automatyzacja ma usuwać pracę jałową, a nie wiedzę instytucjonalną. Pracownicy przechodzą z przepisywania, dekretacji i ręcznych zestawień do analizy, relacji, kontroli jakości, obsługi wyjątków, audytu, bezpieczeństwa i projektowania usług.

Nowa rola	Odpowiedzialność
AI Service Owner	wynik usługi, budżet, ryzyko, relacja z obywatelem i para sprzężona.
Agent Supervisor	nadzór nad flotą agentów, konfiguracja warstwy kontrolnej, obsługa eskalacji.
Model Validator	niezależne testy jakości, stronniczości, odporności i dryfu.
Data Steward	jakość, metadane, prawa dostępu, proveniencja i log dostępu obywatela.
Algorithmic Auditor	audyt modelu, procesu, logów i skutków społecznych.
Prompt/Knowledge Librarian	kanoniczne prompty, korpus, wersje i źródła.
Machine Interface Steward	maszynowo czytelne zobowiązania, schematy usług i API dla agentów obywateli.
Human Oversight Officer	organizacja znaczącego nadzoru i certyfikacja operatorów.
Accessibility Lead	prosty język, interfejs głosowy, kanał analogowy i testy z użytkownikami.
Decommission Manager	wyłączenie systemu, migracja, los danych i spraw w toku.

10.3. Program 2G: tempo i pamięć



Rysunek 12. Model dual-key łączący Gen Z i seniorów w administracji AI.

Model 2G jest wpisany w strukturę zespołów, a nie pozostawiony jako hasło. Młody lider technologiczny odpowiada za tempo, narzędzia i eksperyment. Doświadczony mentor odpowiada za konsekwencje, zgodność, pamięć i ciągłość. Rozbieżność jest dokumentowana i rozstrzygana przez właściciela wyniku. W wymiarze obywatelskim Program 2G działa symetrycznie: mentorzy cyfrowi pomagają seniorom korzystać z agentów obywatelskich, a seniorzy uczą młodych rozpoznawania manipulacji i ryzyka.

Podstawa: destylat NAI I.5–6, III.10, IV.12; wydanie pierwsze „Narzędzia AI”, rozdz. 11 i 15; TS-2G Program 2G; polityka kadrowa i senioralna TechnoRepubliki.

11. Pilotaże, KPI, KRI i rozliczalność

Pilotaż ma dostarczyć dowód, nie prezentację

11.1. Bramy wdrożeniowe

Każdy pilotaż rozwiązuje konkretny problem, ma właściciela, baseline, grupę porównawczą, budżet całego cyklu życia, plan wyjścia, parę sprzężoną po stronie obywatela oraz publiczny raport. Sandbox nie służy obchodzeniu prawa. Służy kontrolowanemu eksperymentowi z jasno zdefiniowanymi granicami, użytkownikami, danymi i warunkiem zatrzymania. Polska ustawa o systemach sztucznej inteligencji tworzy piaskownice pod nadzorem KRIbsi ze ścieżką regulacyjną i technologiczną — program TechnoRepubliki proponuje, aby każda z nich testowała pary narzędzi, a nie pojedyncze systemy.

Brama	Pytanie decyzyjne	Dowód wymagany
G0 — potrzeba	czy problem jest realny i ważniejszy niż alternatywy?	mapa procesu, koszt obecnego stanu, właściciel wyniku
G1 — gotowość	czy istnieją dane, prawo, bezpieczeństwo, para obywatelska i zdolność utrzymania?	assessment, DPIA/FRIA, plan ciągłości, specyfikacja interfejsu dla agentów
G2 — sandbox	czy system i jego para działają w kontrolowanym środowisku?	testy jakości, red team, dostępność, warstwa kontrolna, logi
G3 — pilot	czy poprawia wynik w porównaniu z baseline'em?	telemetria, grupa kontrolna, błędy, koszt, odwołania, obciążenie
G4 — produkcja	czy można skalować bez utraty praw i suwerenności?	SLA, audyt, interoperacyjność, plan wyjścia
G5 — przegląd	czy korzyść utrzymuje się po zmianie danych i kontekstu?	monitoring, dryf, incydenty, test regresji wartości
G6 — dekomisja	czy system nadal ma mandat i sens?	rekord zamknięcia, migracja, los logów i spraw

Pierwsze pilotaże dotyczą: klasyfikacji dokumentów, sprawdzania kompletności, asystenta prawa z korpusem kanonicznym, inteligentnych formularzy z interfejsem dla agentów, anonimizacji, monitoringu terminów, AIOps w trybie cienia, analizy umów oraz — po stronie obywatela — one-click DSAR, logu dostępu do danych i obywatelskiego czytnika rejestru algorytmów.

11.2. Pulpit wskaźników

Pulpit mierzy cztery poziomy: agenta, proces, usługę i sterowność całego systemu, a od drugiego wydania także piąty — symetrię. Nie wystarczy liczba wdrożeń ani osób przeszkolonych.

Typ	Wskaźnik	Interpretacja
KPI	czas sprawy end-to-end	czy obywatel otrzymuje rezultat szybciej
KPI	first-time-right i liczba wezwań do uzupełnienia	czy proces jest prostszy, a dane lepsze
KPI	telemetryczna oszczędność czasu	czy AI rzeczywiście odciąża
KPI	odsetek decyzji z Kwitem Decyzji	czy system jest wyjaśnialny i zaskarżalny
KPI	wskaźnik symetrii: odsetek systemów administracji z działającą parą obywatelską	czy architektura jest symetryczna w praktyce
KPI	odsetek usług z maszynowo czytelnym interfejsem dla agentów	czy agent obywatela nie musi zgadywać
KPI	czas realizacji one-click DSAR	czy obywatel faktycznie kontroluje swoje dane
KRI	odstąpienia od rekomendacji poniżej 2%	ryzyko rubber-stampingu i pozornego nadzoru
KRI	nagły wzrost wniosków lub eskalacji	ryzyko zalewu, ataku lub dryfu

Typ	Wskaźnik	Interpretacja
KRI	odsetek propozycji agentów zablokowanych bez naprawy	ryzyko paraliżu warstwy kontrolnej
KRI	brak aktualizacji rejestru po zmianie wersji	utrata proveniencji decyzji
KRI	wyższy wskaźnik odrzuceń pism przygotowanych z pomocą AI	ryzyko dyskryminacji narzędzi obywatelskich
KCI	odsetek systemów z działającym monitoringiem, audytem i testem regresji wartości	czy bezpieczniki są realne, nie papierowe

Niezależny audyt obejmuje dane, model, proces, interfejs, logi, nadzór człowieka, skutki dla grup oraz zgodność z prawem. Audytor nie używa tego samego modelu bazowego i tych samych danych co agent wykonawczy bez dodatkowej kontroli. Wyniki istotne dla praw obywateli są publiczne w zrozumiałym streszczeniu i w formacie maszynowym.

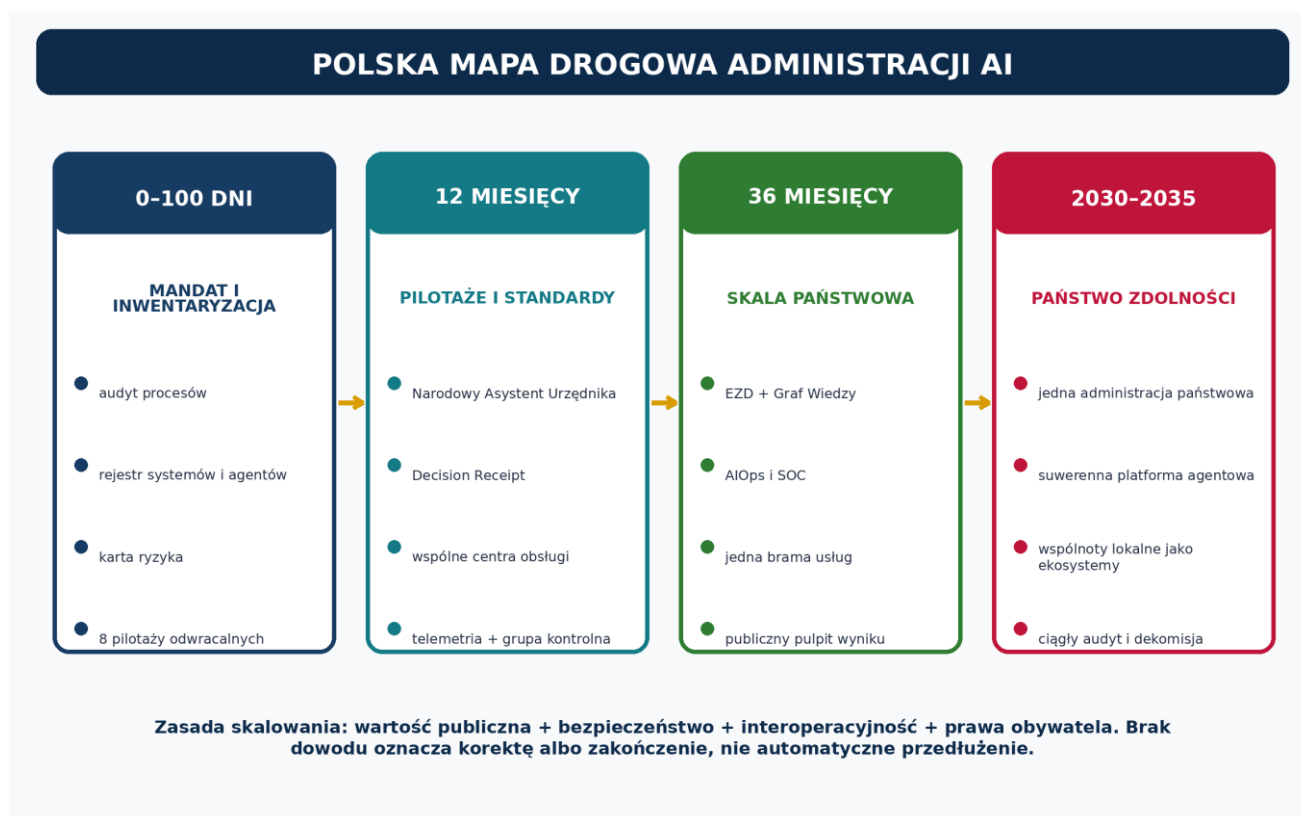
PUBLICZNY REJESTR

Obywatel i jego agent widzą: cel, podstawę prawną, właściciela, dane, wersję modelu, poziom autonomii, parę sprzężoną, znane ograniczenia, datę audytu, liczbę odwołań i status systemu.

Podstawa: TS-P roadmapa i zasada skalowania; TS-SB Sandbox regulacyjny; TS-CC Controlling cybernetyczny; ustawa o systemach sztucznej inteligencji (2026); destylat NAI I.2, V.1, V.5.

12. Mapa drogowa 2026–2035

Od inwentaryzacji do jednolitej, suwerennej i symetrycznej administracji



Rysunek 13. Cztery horyzonty transformacji administracji AI TechnoRepubliki.

Transformacja przebiega w dwóch równoległych torach. Tor operacyjny może rozpocząć się w obecnych kompetencjach i w ramach obowiązującego prawa: inwentaryzacja, polityka dopuszczalnego użycia, standardy interfejsów, pilotaże par sprzężonych, rejestr, szkolenia, EZD, RAG i AIOps. Tor ustrojowy wymaga debaty konstytucyjnej nad jednolitą administracją państwową, nową rolą wspólnot lokalnych i konstytucyjnymi prawami obywatela w epoce agentów.

Horyzont	Decyzje
0–100 dni	narodowy audyt procesów, systemów i agentów AI; zakaz shadow AI na danych wrażliwych; Karta Agenta i Karta Pary Sprzężonej; wybór pilotaży; standard Kwitu Decyzji; zespół 2G; przygotowanie do obowiązków art. 50 AI Act i trybu skargowego KRiBSI.
12 miesięcy	rejestr algorytmów publicznych z API; Narodowy Asystent Urzędnika; warstwa kontrolna agentów w pierwszych resortach; standard maszynowo czytelnych zobowiązań; one-click DSAR i log dostępu w mObywatelu; portfel EUDI w użyciu; piaskownice par sprzężonych; baseline i telemetria.
36 miesięcy	integracja EZD, rejestrów i Grafu Wiedzy z publicznym API; państwowa platforma agentowa z kontrolą A2A; jedna brama usług dla ludzi i agentów; publiczne narzędzia samoobrony prawnej i audytu; personhood credential w konsultacjach; pełna zgodność z obowiązkami dla systemów wysokiego ryzyka (grudzień 2027); publiczny pulpit KPI/KRI.
2030–2035	jednolita administracja państwowa; terenowe centra obsługi; pełna interoperacyjność; suwerenny wariant krytycznych usług; Public Constitutional AI jako stały proces; wspólnoty lokalne jako ekosystemy misji i przedsiębiorczości.

Ustawowy pakiet przejściowy obejmuje: ustawę o administracji AI i parach sprzężonych, nowelizację ustawy o systemach sztucznej inteligencji w zakresie rady społecznej i rejestru algorytmów, standard algorytmicznego due process, prawo do własnego agenta i do maszynowego interfejsu, nowelizację prawa zamówień, ustawę o Narodowym Grafie Wiedzy, ustawę o terenowej administracji państwowej i wspólnotach lokalnych oraz przepisy kadrowe chroniące ciągłość służby publicznej.

DECYZJA USTROJOWA

Lokalna aktywność nie potrzebuje lokalnej biurokracji, lecz praw, zasobów, narzędzi i szybkiej administracji. Państwo zapewnia jednolite reguły i interfejsy; obywatel i wspólnota tworzą wartość z pomocą własnych agentów.

Podstawa: TS-P dziewięć postulatów; projekt Konstytucji TechnoRepubliki Polskiej; PL-CON legalna ścieżka zmiany ustroju; EU-AI harmonogram stosowania; eIDAS 2.0.

13. Test zgodności z formułą, napięcia i warunek falsyfikacji

Formuła, której nie da się obalić, nie jest programem, lecz deklaracją

13.1. Test zgodności rozdziałów z formułą naczelną

Rozdział	Rewolucja agentowa	Symetria	Sprzężenie	Mandat człowieka
3. Architektura symetryczna	●	●	●	●
4. Autonomia i nadzór	●	○	●	●
5. Narzędzia administracji	●	●	●	●
6. Narzędzia obywatela	●	●	●	●
7. Prawa i Public Constitutional AI	○	●	●	●
8. Zarządzanie modelami	●	●	●	●
9. Bezpieczeństwo i suwerenność	●	○	●	●
10. Gotowość i kadry	●	●	○	●
11. Pilotáže i wskaźniki	●	●	●	●
12. Mapa drogowa	●	●	●	●

● — człon formuły rozwinięty wprost; ○ — człon obecny pośrednio. Żaden rozdział nie pozostaje bez powiązania z co najmniej trzema członami formuły.

13.2. Napięcia nierozstrzygnięte

Napięcie	Stan w dokumencie
once-only a ryzyko nadzoru	rozstrzygnięte warunkowo: wymiana rozproszona z logiem dostępu; wymaga oceny skutków dla każdego rejestru
prawo do własnego agenta a pojemność usług	nierozstrzygnięte: wymaga pomiaru zalewu w pilotażach i wyceny reformy pojemności
personhood credential a anonimowość polityczna	rozstrzygnięte kierunkowo: dowód unikalności bez ujawniania tożsamości; kryptografia wymaga audytu
wyjaśnialność a atrybucja odpowiedzialności	nierozstrzygnięte: prawo UE nie różnicuje kształtów HITL; potrzebna regulacja krajowa
suwerenność a jakość modeli granicznych	rozstrzygnięte: zarządzana współzależność i reguła najstabszej warstwy
rules-first a elastyczność uznania administracyjnego	rozstrzygnięte warunkowo: reguły dla rozstrzygnięcia, człowiek dla wyjątku
agent jako pełnomocnik polityczny	rozstrzygnięte: zakres doradczy i wykonawczy, bez głosowania ustrojowego
adaptacja wzorców zagranicznych do Polski	nierozstrzygnięte: mało polskich studiów przypadku; wymaga pracy prawno-instytucjonalnej
jednolita administracja a konstytucyjny samorząd	nierozstrzygnięte w prawie obowiązującym: wymaga zmiany Konstytucji i referendum

13.3. Warunek falsyfikacji

WARUNEK FALSYFIKACJI

Formuła naczelna byłaby obalona, gdyby dało się wykazać, że państwo wdrażające agentów AI wyłącznie po stronie administracji — bez równoważnych narzędzi obywatela, bez sprzężenia przez wspólne zobowiązania i bez znaczącego nadzoru człowieka — trwale zwiększa zarówno sprawność, jak i realną podmiotowość obywateli, mierzoną skutecznością odwołań, kontrolą nad danymi i udziałem w stanowieniu reguł.

Materiał destylatu wskazuje zależność odwrotną: rejestry bez odbiorców stają się teatrem, agenci bez warstwy kontrolnej wykonują niebezpieczne czynności, narzędzia obywatelskie bez zobowiązań państwa są jałowe, a systemy rozstrzygające bez reguł jawnych prowadzą do masowej krzywdy.

13.4. Pięć warunków, przy których architekturę należałoby zmienić

1. gdyby w pilotażach wskaźnik symetrii rósł, a skuteczność odwołań i kontrola nad danymi nie zmieniały się przez co najmniej dwa cykle;
2. gdyby koszt utrzymania par sprzężonych przekraczał wartość korygowanych błędów i oszczędzonego czasu obywateli;
3. gdyby warstwa kontrolna agentów trwale obniżała skuteczność usług poniżej poziomu sprzed wdrożenia mimo ścieżki naprawy;
4. gdyby personhood credential okazał się mechanizmem wykluczenia, którego nie da się usunąć kanałem alternatywnym;
5. gdyby rules-first systematycznie dawało gorsze rozstrzygnięcia w sprawach wyjątkowych niż nadzorowane modele probabilistyczne, przy porównywalnej rozliczalności.

REKOMENDACJE KOŃCOWE — 12 DECYZJI

1. Ustanowić zasadę pary sprzężonej: żadne narzędzie AI administracji bez narzędzia obywatela.
2. Zbudować warstwę zerową: personhood credential, portfel EUDI i tożsamość maszynową agentów.
3. Wprowadzić prawo obywatela do własnego agenta i obowiązek maszynowo czytelnych zobowiązań państwa.
4. Objąć wszystkich agentów administracji warstwą kontrolną z naprawą propozycji i kontrolą komunikacji A2A.
5. Automatyzować według odwracalności, a decyzje prawnie doniosłe rozstrzygać regułami jawnymi.
6. Zdefiniować human-in-command przez pojemność, kompetencje, prawo zatrzymania i symetrię uzasadnienia.
7. Wprowadzić Kwit Decyzji AI, algorytmiczny due process i one-click DSAR z logiem dostępu.
8. Przekształcić rejestr algorytmów w obowiązkową, ustrukturyzowaną bramkę z API i obywatelskim toolkitem audytu.
9. Zbudować bi-temporalny Narodowy Graf Wiedzy z publicznym dostępem i zakazem cytowań spoza korpusu kanonicznego.
10. Uruchomić Public Constitutional AI: współtworzenie, kontestację i widoczne egzekwowanie zasad AI państwa.
11. Prowadzić suwerenność jako zarządzaną współzależność w trzech wymiarach i finansować publiczną infrastrukturę narzędzi obywatelskich.
12. Skalować wyłącznie pary narzędzi, które udowodniły wartość publiczną, bezpieczeństwo i zgodność z prawami obywatela, łącząc wdrożenia z reformą organizacyjną i Programem 2G.

ZDANIE PROGRAMOWE

O narzędziach AI TechnoRepubliki rozstrzyga nie to, czy państwo ma agentów, lecz czy obywatel ma równie skuteczne narzędzia, by państwo zrozumieć, wykorzystać, zakwestionować i skorygować — oraz czy człowiek po obu stronach ma wiedzę, czas i uprawnienie, aby zatrzymać system.

ANEKS A. MACIERZ KLASYFIKACJI ZASTOSOWANIA AI

Pytanie	0 — niskie	1 — średnie	2 — wysokie	Uwagi
Skutek prawny	brak	odwracalny lub pośredni	bezpośredni i trudny do odwrócenia	rules-first przy wartości 2
Uznaniowość	reguły jednoznaczne	część wyjątków	ocena indywidualna i wartościująca	blind review przy 2
Dane	jawne lub techniczne	dane osobowe	dane wrażliwe / profilowanie	DPIA przy 1 i 2
Odwracalność	natychmiastowy rollback	korekta w SLA	skutek trwały lub represyjny	zakaz A3+ przy 2
Walidacja	wynik obiektywny	wielokryterialna	brak pełnego ground truth	AHP z CR przy 1
Zasięg	wewnętrzny	grupa obywateli	masowy lub infrastruktura krytyczna	CER i NIS2 przy 2
Autonomia agenta	brak czynności zewnętrznych	czynności odwracalne	czynności wobec obywatela lub rejestru	warstwa kontrolna obowiązkowa przy 1 i 2

Suma punktów 0–3: A0–A2. Suma 4–8: co najmniej A1 z bramkami i warstwą kontrolną. Suma 9–14: system wysokiego wpływu — human-in-command, FRIA/DPIA, audyt niezależny, para sprzężona i ograniczenie autonomii. Zastosowania A5 pozostają zakazane niezależnie od wyniku.

ANEKS B. MINIMALNA KARTA AGENTA / SYSTEMU AI

Pole	Wymagana treść
Identyfikator	nazwa, wersja, hash/ID, status: test/produkcja/zawieszony/wycofany.
Właściciel	organ, AI Service Owner, kontakt, odpowiedzialność za wynik.
Mandat	podstawa prawna, cel, dozwolone i zakazane zastosowania.
Para sprzężona	narzędzie obywatela, interfejs maszynowy, format uzasadnienia.
Poziom autonomii	A0–A5, odwracalność, punkty zatwierdzenia i ścieżka eskalacji.
Tożsamość maszynowa	identyfikator agenta, dostawca poświadczeń, czas życia tokenów.
Dane	źródła, minimalizacja, jakość, retencja, przestrzeń danych, lineage.
Narzędzia	lista dozwolonych narzędzi, uprawnienia, limity budżetu i tempa.
Komunikacja A2A	dozwoleni partnerzy, propagacja mandatu, atestacja zdolności.

Pole	Wymagana treść
Warstwa kontrolna	reguły kwalifikacji Approve/Revise/Block/Escalate, dziennik.
Model	dostawca, wersja, hosting, benchmark, znane ograniczenia.
Nadzór	role, kształt HITL/HOTL/HIC, limit obciążenia, blind review, zasada czterech oczu.
Logi i audyt	zakres logowania, retencja, dostęp audytora i obywatela, raporty.
Stop/dekomisja	wyłącznik awaryjny, rollback, warunek zatrzymania, plan migracji.

ANEKS C. KWIT DECYZJI AI

Sekcja kwitu	Treść dla obywatela
Kto odpowiada	organ, osoba lub rola zatwierdzająca, kontakt.
Jakie AI uczestniczyło	nazwa, wersja, rola: informacja/rekomendacja/wykonanie; odnośnik do rejestru.
Ścieżka warstwy kontrolnej	czy czynność była zatwierdzona, poprawiona lub eskalowana.
Input Snapshot	dane i dokumenty użyte w sprawie, data pobrania, możliwość korekty.
Podstawa prawna	akty prawne i przepisy obowiązujące w dniu decyzji (stan bi-temporalny).
Wpływ na wynik	najważniejsze czynniki i kierunek ich wpływu.
Kontrfaktyk	jaka zmiana danych lub warunków mogłaby zmienić wynik.
Niepewność	ograniczenia, brakujące dane, poziom pewności.
Odwołanie	procedura, termin, kanał ludzki, numer sprawy; format maszynowy dla agenta obywatela.

ANEKS D. KARTA PILOTAŻU I RAPORT PUBLICZNY

Element	Pytanie kontrolne
Problem i baseline	jaki wynik, czas, koszt i błąd występują bez AI?
Hipoteza	co dokładnie ma się poprawić i o ile?
Grupa porównawcza	jak odróżnimy efekt systemu od innych zmian?
Para sprzężona	jaki narzędzie obywatela jest testowane równolegle i jak mierzymy jego skuteczność?
Prawa i ryzyka	kogo system dotyka, jaki jest najgorszy realistyczny skutek?
Dane	czy są autorytatywne, reprezentatywne i aktualne?
Nadzór	kto może zatrzymać i odwrócić działanie?
Koszt całego cyklu	wdrożenie, utrzymanie, audyt, migracja i dekomisja.

Element	Pytanie kontrolne
Warunek skalowania	jaki dowód otwiera kolejną bramę?
Warunek zatrzymania	jaki próg błędu, kosztu, zalewu lub naruszenia zamyka pilot?
Raport	co publikujemy: efekt, błędy, incydenty, odwołania i decyzję końcową?

ANEKS E. KARTA PARY SPRZĘŻONEJ

Karta jest nowym instrumentem drugiego wydania. Wypełnia się ją przed bramą G1 dla każdego narzędzia AI administracji.

Pole	Wymagana treść
Narzędzie administracji	nazwa, karta systemu, poziom autonomii.
Narzędzie obywatela	nazwa, dostawca publiczny, kanały dostępu (aplikacja, głos, analogowy).
Asymetria wyjściowa	co państwo wie lub może zrobić, czego obywatel nie wie lub nie może.
Zobowiązanie maszynowe	schemat danych, endpoint, termin, format uzasadnienia.
Warstwa zerowa	wymagane poświadczenia, delegacja, dowód unikalności.
Kanał zwrotny	skarga, odwołanie, sygnał; instytucja z obowiązkiem reakcji.
Miernik symetrii	jak mierzymy, że obywatel realnie korzysta z pary.
Ryzyko zalewu	prognoza obciążenia i mechanizm pojemności.
Test inkluzji	wyniki dla seniorów, osób z niepełnosprawnościami i osób wykluczonych cyfrowo.
Warunek wstrzymania	przy jakiej asymetrii narzędzie administracji zostaje zawieszone.

ANEKS F. SŁOWNIK POJĘĆ, TERMINÓW I SKRÓTÓW

F.1. Skróty, akronimy i oznaczenia

Skrót	Rozwinięcie i znaczenie
A0–A5	skala poziomów autonomii systemu AI: informacja, rekomendacja, rutyna, delegacja, adaptacja, władza strategiczna (A5 zakazane).
A2A	Agent2Agent — protokół komunikacji między agentami AI różnych dostawców i organizacji.
AHP	Analytic Hierarchy Process — analityczny proces hierarchiczny; metoda wielokryterialna oparta na porównaniach parami (Saaty).
AI	sztuczna inteligencja (Artificial Intelligence).
AI Act	rozporządzenie (UE) 2024/1689 w sprawie sztucznej inteligencji.
AIOps	wykorzystanie AI w utrzymaniu i eksploatacji systemów informatycznych.

Skrót	Rozwinięcie i znaczenie
API	Application Programming Interface — maszynowy interfejs programistyczny.
CER	dyrektywa (UE) 2022/2557 w sprawie odporności podmiotów krytycznych.
CR / CI	Consistency Ratio / Consistency Index — wskaźniki spójności porównań w AHP (próg $CR \leq 0,10$).
DDoS	rozproszony atak odmowy usługi.
DLP	Data Loss Prevention — zapobieganie wyciekom danych.
DPI	Digital Public Infrastructure — cyfrowa infrastruktura publiczna: tożsamość, płatności, wymiana danych.
DPIA	ocena skutków dla ochrony danych (art. 35 RODO).
DSAR	Data Subject Access Request — żądanie dostępu osoby do jej danych.
DSS	Decision Support System — system wsparcia decyzji.
eIDAS 2.0	rozporządzenie (UE) 2024/1183 zmieniające eIDAS, ustanawiające europejskie ramy tożsamości cyfrowej.
EUDI Wallet	European Digital Identity Wallet — Europejski Portfel Tożsamości Cyfrowej.
EZD	elektroniczne zarządzanie dokumentacją.
FRIA	Fundamental Rights Impact Assessment — ocena skutków dla praw podstawowych (art. 27 AI Act).
GPAI	General-Purpose AI — modele AI ogólnego przeznaczenia.
GraphRAG	RAG wykorzystujący graf wiedzy.
HIC	Human-in-Command — człowiek decyduje i dowodzi systemem.
HITL	Human-in-the-Loop — człowiek zatwierdza lub uczestniczy w pętli decyzji.
HOTL	Human-on-the-Loop — człowiek nadzoruje system i może interweniować.
ISM	Interpretive Structural Modelling — modelowanie strukturalne zależności kryteriów.
KCI	Key Control Indicator — kluczowy wskaźnik kontroli.
KPI	Key Performance Indicator — kluczowy wskaźnik wyniku.
KRI	Key Risk Indicator — kluczowy wskaźnik ryzyka.
KRiBSI	Komisja Rozwoju i Bezpieczeństwa Sztucznej Inteligencji — polski organ nadzoru rynku AI.
LLM	Large Language Model — duży model językowy.
MCDM	Multi-Criteria Decision Making — wielokryterialne podejmowanie decyzji.
MCP	Model Context Protocol — otwarty protokół łączenia modeli i agentów z narzędziami i danymi.
NIS2	dyrektywa (UE) 2022/2555 w sprawie cyberbezpieczeństwa.

Skrót	Rozwinięcie i znaczenie
OCL	Organizational Control Layer — warstwa kontrolna nad agentami organizacji.
OIDC	OpenID Connect — protokół uwierzytelniania oparty na OAuth 2.0.
PDV	Personal Data Vault — osobisty skarbiec danych.
PLLuM	Polish Large Language Model — rodzina polskich otwartych modeli językowych.
PUODO	Prezes Urzędu Ochrony Danych Osobowych.
RAD	Retrieval-Augmented Decision-making — decyzje wspomagane wyszukiwaniem i AHP.
RAG	Retrieval-Augmented Generation — generowanie odpowiedzi wspomagane wyszukiwaniem w korpusie.
RODO	rozporządzenie (UE) 2016/679 o ochronie danych osobowych.
S1–S5, S3*	funkcje modelu systemu żywotnego: operacje, koordynacja, kontrola, audyt, rozpoznanie przyszłości, tożsamość.
SLA	Service Level Agreement — gwarantowany poziom usługi.
SSI	Self-Sovereign Identity — samосуwerenna tożsamość.
TK	Trybunał Konstytucyjny.
TSUE	Trybunał Sprawiedliwości Unii Europejskiej.
VC	Verifiable Credentials — weryfikowalne poświadczenia.
VSM	Viable System Model — model systemu żywotnego (Beer).
WCAG	Web Content Accessibility Guidelines — wytyczne dostępności treści internetowych.
XAI	Explainable AI — wyjaśnialna sztuczna inteligencja.
ZKP	Zero-Knowledge Proof — dowód o wiedzy zerowej.

F.2. Pojęcia sztucznej inteligencji i systemów agentowych

Pojęcie	Definicja
Agent AI	system wykorzystujący model AI do samodzielnego planowania i wykonywania wieloetapowych zadań z użyciem narzędzi, pamięci i uprawnień, w granicach mandatu.
Agent obywatelski	agent działający na podstawie jawnej, odwołalnej delegacji obywatela w kontaktach z państwem; zakres doradczy i wykonawczy, bez głosowania ustrojowego.
Agentic flooding	przeciążenie usług publicznych wnioskami, pismami i zapytaniami generowanymi automatycznie przez agentów.
Agentic inequality	nierówność wynikająca z różnej jakości agentów dostępnych dla różnych grup społecznych.

Pojęcie	Definicja
Advance Control Directive	dyrektywa wyprzedzająca: ustanowiony z góry zakres działań systemu dla decyzji poza czasem ludzkiej percepcji; odpowiedzialność przypisana momentowi ustanowienia.
Automation bias	skłonność człowieka do nadmiernego polegania na wyniku systemu automatycznego.
Bi-temporalność	zapis dwóch osi czasu faktu: czasu obowiązywania w świecie i czasu, w którym system o nim wiedział.
Deterministyczny walidator	komponent regułowy sprawdzający wynik modelu (np. każde cytowanie w kanonicznym korpusie).
Dryf modelu	pogorszenie jakości lub zmiana zachowania modelu wskutek zmiany danych, kontekstu lub wersji.
Horyzont zadań	długość zadania (mierzona czasem pracy eksperta), które agent kończy z zadaną skutecznością.
Korpus kanoniczny	autorytatywny, wersjonowany zbiór źródeł prawa i instrukcji, jedyne dopuszczalne źródło cytowań.
Kwit Decyzji	ustandaryzowany dokument dla obywatela opisujący rolę AI, dane, podstawę prawną, czynniki wpływu, kontrfaktyk i drogę odwołania.
LLM-as-a-judge	wykorzystanie modelu językowego jako oceniającego jakość odpowiedzi lub wariantów.
Maszynowo czytelne zobowiązania	ustandaryzowane, formalnie typowane interfejsy usług publicznych (wniosek, status, uzasadnienie, termin, DSAR) dostępne dla agentów obywateli.
Para sprzężona	narzędzie AI administracji i odpowiadające mu narzędzie AI obywatela, połączone wspólnym zobowiązaniem i kanałem zwrotnym.
Personhood credential	kryptograficzne poświadczenie, że użytkownik jest unikalnym człowiekiem, bez ujawniania jego tożsamości.
Prompt injection	atak polegający na wprowadzeniu do danych instrukcji przejmujących sterowanie modelem.
Public Constitutional AI	proces legitymizacji zasad AI państwa przez współtworzenie reguł, mechanizm kontestacji i widoczne egzekwowanie.
Rozumowanie w czasie wnioskowania	przeznaczanie przez model dodatkowych obliczeń na analizę przed udzieleniem odpowiedzi.
Rules-first	zasada projektowa: rozstrzygnięcie prawnie doniosłe wynika z reguły jawnej, a model jedynie wspiera.
Shadow AI	używanie niezatwierdzonych narzędzi AI na danych służbowych.
Structured recovery	ustrukturyzowana naprawa propozycji agenta przez warstwę kontrolną zamiast samego blokowania.
Teatr przejrzystości	pozorna jawność: publikacja informacji bez struktury, kompletności, wymuszalności i krytycznego odbiorcy.

Pojęcie	Definicja
Tożsamość maszynowa	identyfikator i poświadczenia agenta, powiązane z mandatem i właścicielem.
Verifier-as-a-Service	usługa orkestracji weryfikacji poświadczeń w wielu formatach dla wielu organów.

F.3. Pojęcia cybernetyczne

Pojęcie	Definicja
Cybernetyka	nauka o sterowaniu i komunikacji w organizmach, maszynach i organizacjach (Wiener).
Cybernetyka drugiego rzędu	cybernetyka systemów obserwujących, w której obserwator jest częścią systemu (von Foerster).
Dryf wartości	stopniowe odchodzenie systemu doradczego od zakotwiczonych wartości ku konwencji.
Kanał algedoniczny	kanal sygnału alarmowego omijający hierarchię i docierający bezpośrednio do poziomu decyzji.
Model systemu żywotnego (VSM)	model pięciu funkcji niezbędnych do przetrwania organizacji: S1 operacje, S2 koordynacja, S3 kontrola (S3* audyt), S4 przyszłość, S5 tożsamość.
Paradoks Salomona	niezdolność instytucji do rozpoznania wiedzy, którą już posiada.
Prawo wymaganej różnorodności	zasada Ashby'ego: tylko różnorodność regulatora może pochłoniąć różnorodność zakłóceń.
Regulator	element systemu utrzymujący wybrane zmienne w dopuszczalnych granicach.
Sprzężenie zwrotne	przekazanie informacji o wyniku działania z powrotem do jego źródła w celu korekty.
Transduktor	element przekładający sygnał jednego systemu na język drugiego.
Twierdzenie o dobrym regulatorze	każdy dobry regulator systemu musi być modelem tego systemu (Conant, Ashby).

F.4. Pojęcia systemowe i sieciowe

Pojęcie	Definicja
Atak Sybil	tworzenie wielu fałszywych tożsamości w celu uzyskania nieproporcjonalnego wpływu w sieci.
Awaria normalna	awaria nieunikniona w systemie złożonym o ścisłych sprzężeniach (Perrow).
Efekt sieciowy	wzrost wartości sieci wraz z liczbą jej uczestników.
Federacja agentów	sieć wyspecjalizowanych agentów za wspólnym punktem wejścia, bez centralnego monolitu.

Pojęcie	Definicja
Granica systemu	decyzja o tym, co należy do analizowanego systemu, przesądzająca o mierze sukcesu.
Hub	węzeł sieci o bardzo wysokiej liczbie połączeń.
Kaskada awarii	propagacja awarii między współzależnymi sieciami lub warstwami.
Model klepsydry	architektura z wąską szyjką wspólnych standardów podtrzymującą wiele aplikacji po obu stronach.
Once-only	zasada jednokrotnego przekazywania danych administracji.
Prawo Goodharta	miara, która staje się celem, przestaje być dobrą miarą.
Punkt dźwigni	miejsce w systemie, w którym niewielka zmiana wywołuje trwały skutek (Meadows).
Sieć bezskalowa	sieć o rozkładzie stopni węzłów zgodnym z prawem potęgowym, z nielicznymi hubami (Barabási, Albert).
Zarządzana współzależność	strategia suwerenności oparta na mapowaniu zależności, priorytetyzacji przewag i dywersyfikacji.

F.5. Pojęcia prawne, instytucjonalne i programowe

Pojęcie	Definicja
Algorytmiczny due process	zespół gwarancji: wyjaśnienie, zakwestionowanie, przegląd człowieka, niezależna kontrola, naprawa szkody.
Dobra administracja	prawo z art. 41 Karty Praw Podstawowych UE: bezstronność, rzetelność, terminowość, wysłuchanie, dostęp do akt, uzasadnienie.
Human-in-command	zdolność człowieka do zrozumienia, zatrzymania i skorygowania systemu, a nie formalny podpis.
Model 3A	metoda TechnoSanacji: Antycypacja, Absorpcja, Adaptacja.
Piaskownica regulacyjna	kontrolowane środowisko testowania systemów AI pod nadzorem organu.
Program 2G	model współpracy Generacji Z i seniorów (dual-key): tempo i narzędzia oraz pamięć i odpowiedzialność.
Rejestr algorytmów publicznych	obowiązkowy, ustrukturyzowany i wersjonowany wykaz systemów AI stosowanych przez organy publiczne.
System wysokiego ryzyka	system AI wymieniony w Załączniku I lub III AI Act, podlegający wymogom jakości, nadzoru i dokumentacji.
TechnoRepublika	cel programu: państwo, które przewiduje, działa i odpowiada przed obywatelem.
TechnoSanacja	droga programowa i ruch obywatelski budujący TechnoRepublikę.
Zdarzenie życiowe	sytuacja człowieka (narodziny, praca, firma, choroba, śmierć bliskiej osoby), wokół której projektuje się usługę.

ANEKS G. BIBLIOGRAFIA

Bibliografia obejmuje źródła przywołane w publikacji oraz podstawowe prace, na których opierają się cztery drogi dowodowe. Liczby i ustalenia z destylatu operacyjnego należy interpretować zgodnie z jego zastrzeżeniami metodologicznymi: pochodzą z próbek reprezentatywnych korpusu i w części wymagają dalszej weryfikacji, a wzorce zagraniczne wymagają adaptacji do polskiego porządku prawnego.

G.1. Dokumenty programowe TechnoSanacji

- [TS-01] TechnoSanacja (2026), Manifest TechnoSanacji, Polska, 14.07.2026, <https://technosanacja.pl/>
- [TS-K] TechnoSanacja (2026), Projekt Konstytucji TechnoRepubliki Polskiej, dokument 05, <https://technosanacja.pl/>
- [TS-P] TechnoSanacja (2026), Postulaty — dziewięć postulatów ustrojowych, <https://technosanacja.pl/postulat/>
- [TS-S] TechnoSanacja (2026), Struktura TechnoRepubliki Polskiej, <https://technosanacja.pl/struktura/>
- [TS-NSB] TechnoSanacja (2026), Narodowa Strategia Bezpieczeństwa TechnoRepubliki Polskiej, t. 1, dokument 04.
- [TS-11] TechnoSanacja (2026), Prezydent — zintegrowana egzekutywa, dokument 11.
- [TS-17] TechnoSanacja (2026), Administracja państwowa, dokument 17.
- [TS-CC] TechnoSanacja (2026), Controlling cybernetyczny TechnoRepubliki, dokument 19.
- [TS-SI] TechnoSanacja (2026), System informacyjny TechnoRepubliki Polskiej, dokument 24.
- [TS-2G] TechnoSanacja (2026), Program 2G, dokument 36.
- [TS-REZ] TechnoSanacja (2026), Rezyliencja — odporność państwa, dokument 38.
- [TS-MAS] TechnoSanacja (2026), Graf multi-agentów TechnoRepubliki Polskiej, dokument 39.
- [TS-SB] TechnoSanacja (2026), Sandbox regulacyjny, dokument 43.
- [TS-45] TechnoSanacja (2026), Infrastruktura odporności państwa, dokument 45.
- [TS-46] TechnoSanacja (2026), Analityczny proces wyboru technologii militarnych, dokument 46.
- [NAI-1] Węgrzyn, A. (2026), Narzędzia AI — dla obywateli i dla administracji państwowej, wydanie pierwsze, TechnoSanacja, 08.08.2026.
- [NAI-D] Destylat NAI (2026), Narzędzia AI — dla obywateli i administracji TechnoRepubliki: synteza 1610 publikacji w klastrach GOVTECH, MCDM, CITIZEN, ETHICS i CITIZENTOOLS, biblioteka badawcza TechnoSanacji.
- [AA-D] Administracja AI TechnoRepubliki — research z korpusu 1501 publikacji, lipiec 2026, biblioteka badawcza TechnoSanacji.

G.2. Akty prawne, orzeczenia i dokumenty publiczne

- [PL-CON] Konstytucja Rzeczypospolitej Polskiej z dnia 2 kwietnia 1997 r., Dz.U. 1997 nr 78 poz. 483, <https://www.sejm.gov.pl/prawo/konst/polski/kon1.htm>
- [PL-AI] Ustawa o systemach sztucznej inteligencji, podpisana 24.07.2026, ogłoszona w Dzienniku Ustaw 27.07.2026; proces legislacyjny: https://orka.sejm.gov.pl/proc10.nsf/ustawy/2443_u.htm
- [PL-DIP] Ustawa z dnia 6 września 2001 r. o dostępie do informacji publicznej.
- [PL-USG] Ustawa z dnia 8 marca 1990 r. o samorządzie gminnym, tekst jednolity.
- [TK-K53] Trybunał Konstytucyjny (2018), wyrok z 6 czerwca 2018 r., sygn. K 53/16 (profilowanie osób bezrobotnych).
- [EU-AI] Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie sztucznej inteligencji (AI Act), Dz.Urz. UE L z 12.07.2024, <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- [EU-OMN] Komisja Europejska (2025), wniosek Digital Omnibus on AI z 19.11.2025; porozumienie polityczne z maja 2026 — omówienie: Gibson Dunn (2026), EU AI Act Omnibus Agreement — Postponed High-Risk Deadlines and Other Key Changes, 27.05.2026, <https://www.gibsondunn.com/eu-ai-act-omnibus-agreement-postponed-high-risk-deadlines-and-other-key-changes/>
- [EU-eIDAS] Rozporządzenie (UE) 2024/1183 z dnia 11 kwietnia 2024 r. zmieniające rozporządzenie (UE) nr 910/2014 w odniesieniu do ustanowienia europejskich ram tożsamości cyfrowej.
- [EU-RODO] Rozporządzenie (UE) 2016/679 (RODO), w szczególności art. 22 i 35.
- [EU-KPP] Karta Praw Podstawowych Unii Europejskiej, art. 41 — prawo do dobrej administracji.

- [EU-NIS2] Dyrektywa (UE) 2022/2555 w sprawie środków na rzecz wysokiego wspólnego poziomu cyberbezpieczeństwa (NIS2).
- [EU-CER] Dyrektywa (UE) 2022/2557 w sprawie odporności podmiotów krytycznych (CER).
- [EU-IEA] Rozporządzenie (UE) 2024/903 — Interoperable Europe Act, <https://interoperable-europe.ec.europa.eu/>
- [TSUE-SCHUFA] TSUE (2023), wyrok z 7 grudnia 2023 r., C-634/21, SCHUFA Holding (Scoring).
- [NL-SyRI] Rechtbank Den Haag (2020), wyrok z 5 lutego 2020 r., ECLI:NL:RBDHA:2020:865 (System Risico Indicatie).
- [AU-RD] Royal Commission into the Robodebt Scheme (2023), Report, Commonwealth of Australia, Canberra.
- [CoE] Rada Europy (2024), Konwencja ramowa o sztucznej inteligencji i prawach człowieka, demokracji i państwie prawa, Wilno, 5.09.2024.
- [SG-MGF] IMDA / Ministry of Digital Development and Information (2026), Model AI Governance Framework for Agentic AI, Singapur, styczeń 2026, <https://www.imda.gov.sg/>
- [PL-SCP] Ministerstwo Cyfryzacji (2025), Strategia Cyfryzacji Państwa do 2035 r., <https://www.gov.pl/web/cyfryzacja/strategia-cyfryzacji-panstwa>
- [PL-EUDI] Ministerstwo Cyfryzacji / COI (2026), Europejski portfel tożsamości cyfrowej zmierza do mObywatela, <https://www.gov.pl/web/cyfryzacja/europejski-portfel-tozsamosci-cyfrowej-zmierza-do-mobywatela>
- [PL-PLLuM] Ministerstwo Cyfryzacji (2025), Więcej, szybciej, po polsku — rodzina modeli PLLuM się powiększa, <https://www.gov.pl/web/cyfryzacja/wiecej-szybciej-po-polsku-rodzina-modeli-pllum-sie-powieksza>
- [PL-EZD] Ministerstwo Cyfryzacji / NASK, EZD RP, <https://www.gov.pl/web/ezd-rp/>
- [RP-2026] Nowotnik, N. (2026), Ustawa o AI weszła już w życie. Od października będzie można się na nią poskarżyć, „Rzeczpospolita”, 11.08.2026.

G.3. Cybernetyka, teoria systemów i teoria sieci

- Ashby, W.R. (1956), *An Introduction to Cybernetics*, Chapman & Hall, Londyn.
- Conant, R.C., Ashby, W.R. (1970), Every Good Regulator of a System Must Be a Model of That System, „International Journal of Systems Science” 1(2), s. 89–97.
- Wiener, N. (1948), *Cybernetics: Or Control and Communication in the Animal and the Machine*, MIT Press, Cambridge MA.
- Beer, S. (1972), *Brain of the Firm*, Allen Lane, Londyn.
- Beer, S. (1979), *The Heart of Enterprise*, John Wiley & Sons, Chichester.
- Beer, S. (1985), *Diagnosing the System for Organizations*, John Wiley & Sons, Chichester.
- von Foerster, H. (2003), *Understanding Understanding: Essays on Cybernetics and Cognition*, Springer, Nowy Jork.
- Medina, E. (2011), *Cybernetic Revolutionaries: Technology and Politics in Allende's Chile*, MIT Press, Cambridge MA.
- Meadows, D.H. (1999), *Leverage Points: Places to Intervene in a System*, The Sustainability Institute, Hartland VT.
- Meadows, D.H. (2008), *Thinking in Systems: A Primer*, Chelsea Green, White River Junction.
- Perrow, C. (1984), *Normal Accidents: Living with High-Risk Technologies*, Basic Books, Nowy Jork.
- Strathern, M. (1997), „Improving ratings”: audit in the British University system, „European Review” 5(3), s. 305–321.
- Barabási, A.-L., Albert, R. (1999), Emergence of Scaling in Random Networks, „Science” 286(5439), s. 509–512.
- Buldyrev, S.V., Parshani, R., Paul, G., Stanley, H.E., Havlin, S. (2010), Catastrophic cascade of failures in interdependent networks, „Nature” 464, s. 1025–1028.
- Newman, M.E.J. (2010), *Networks: An Introduction*, Oxford University Press, Oksford.
- Douceur, J.R. (2002), The Sybil Attack, w: *Peer-to-Peer Systems (IPTPS 2002)*, LNCS 2429, Springer, s. 251–260.
- Ostrom, E. (1990), *Governing the Commons*, Cambridge University Press, Cambridge.

G.4. Sztuczna inteligencja, agenci i governance

- Stanford HAI (2025), *Artificial Intelligence Index Report 2025*, Stanford University, <https://hai.stanford.edu/ai-index/2025-ai-index-report>
- Kwa, T., West, B., Becker, J. i in. (2025), *Measuring AI Ability to Complete Long Tasks*, METR, arXiv:2503.14499.
- METR (2026), *Time Horizon 1.1*, 29.01.2026, <https://metr.org/blog/2026-1-29-time-horizon-1-1/>

- Anthropic (2024), Introducing the Model Context Protocol, 25.11.2024; Anthropic (2025), Donating the Model Context Protocol and establishing the Agentic AI Foundation, grudzień 2025, <https://anthropic.com/news/donating-the-model-context-protocol-and-establishing-of-the-agentic-ai-foundation>
- Linux Foundation (2025), Linux Foundation Announces the Formation of the Agentic AI Foundation (AAIF), 9.12.2025, <https://www.linuxfoundation.org/>
- Google (2025), Announcing the Agent2Agent Protocol (A2A), 9.04.2025, <https://developers.googleblog.com/>
- Yao, S. i in. (2023), ReAct: Synergizing Reasoning and Acting in Language Models, ICLR 2023.
- Lewis, P. i in. (2020), Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, „Advances in Neural Information Processing Systems” 33.
- Zheng, L. i in. (2023), Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena, NeurIPS 2023 Datasets and Benchmarks.
- Bai, Y. i in. (2022), Constitutional AI: Harmlessness from AI Feedback, arXiv:2212.08073.
- Abiri, G. (2025), Public Constitutional AI, „Georgia Law Review” 59, s. 601 i n.; arXiv:2406.16696.
- Adler, S., Hitzig, Z., Jain, S. i in. (2024), Personhood credentials: Artificial intelligence and the value of privacy-preserving tools to distinguish who is real online, arXiv:2408.07892.
- Shavit, Y. i in. (2023), Practices for Governing Agentic AI Systems, OpenAI.
- Chan, A. i in. (2024), Visibility into AI Agents, FAccT '24, ACM.
- Parasuraman, R., Riley, V. (1997), Humans and Automation: Use, Misuse, Disuse, Abuse, „Human Factors” 39(2), s. 230–253.
- Green, B. (2022), The flaws of policies requiring human oversight of government algorithms, „Computer Law & Security Review” 45, 105681.
- Wachter, S., Mittelstadt, B., Russell, C. (2018), Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR, „Harvard Journal of Law & Technology” 31(2), s. 841–887.
- Mitchell, M. i in. (2019), Model Cards for Model Reporting, FAT* '19, ACM, s. 220–229.
- Gebri, T. i in. (2021), Datasheets for Datasets, „Communications of the ACM” 64(12), s. 86–92.
- Raji, I.D. i in. (2020), Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing, FAT* '20, ACM, s. 33–44.
- Floridi, L. i in. (2018), AI4People — An Ethical Framework for a Good AI Society, „Minds and Machines” 28, s. 689–707.
- NIST (2023), Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1, Gaithersburg.
- ISO/IEC 42001:2023, Information technology — Artificial intelligence — Management system.
- OWASP (2024), OWASP Top 10 for LLM Applications 2025, OWASP Foundation.
- OECD (2024), Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, zmieniona 2024.
- Snodgrass, R.T. (1999), Developing Time-Oriented Database Applications in SQL, Morgan Kaufmann, San Francisco.

G.5. Administracja publiczna, GovTech i obywatel

- Moynihan, D., Herd, P., Harvey, H. (2015), Administrative Burden: Learning, Psychological, and Compliance Costs in Citizen–State Interactions, „Journal of Public Administration Research and Theory” 25(1), s. 43–69.
- Herd, P., Moynihan, D.P. (2018), Administrative Burden: Policymaking by Other Means, Russell Sage Foundation, Nowy Jork.
- Dunleavy, P., Margetts, H., Bastow, S., Tinkler, J. (2006), Digital Era Governance: IT Corporations, the State, and E-Government, Oxford University Press, Oksford.
- Saaty, T.L. (1980), The Analytic Hierarchy Process, McGraw-Hill, Nowy Jork.
- Koczkodaj, W.W. (1993), A New Definition of Consistency of Pairwise Comparisons, „Mathematical and Computer Modelling” 18(7), s. 79–84.
- World Bank (2022), GovTech Maturity Index 2022 Update: Trends in Public Sector Digital Transformation, Waszyngton.
- World Bank, Identification for Development (ID4D) i Digital Public Infrastructure — materiały koncepcyjne, <https://id4d.worldbank.org/>
- e-Estonia, X-Road i Bürokratt — dokumentacja programów estońskiej administracji cyfrowej, <https://e-estonia.com/>
- Ministerstwo Transformacji Cyfrowej Ukrainy, Diia — portal i aplikacja usług publicznych, <https://diia.gov.ua/>

Carney, T. (2019), Robo-debt illegality: The seven veils of failed guarantees of the rule of law?, „Alternative Law Journal” 44(1), s. 4–10.

Lessig, L. (1999), Code and Other Laws of Cyberspace, Basic Books, Nowy Jork.

G.6. Korpus źródłowy destylatu NAI

Destylat operacyjny NAI powstał z 1610 dokumentów PDF (NAI_GOVTECH 303, NAI_MCDM 159, NAI_CITIZEN 304, NAI_ETHICS 492, NAI_CITIZENTOOLS 352), scalonych do 14 części i destylowanych przez pięć digestów pośrednich. Opisywane w publikacji wyniki badań — m.in. warstwa kontrolna agentów (redukcja niebezpiecznych wykonań z 88% do 0%), systemy JUSTICE i ARGOS, benchmark DisasterBench, wzorzec RAD, symulacja Cybersyn, analiza ośmiu gmin holenderskich, 84 przypadki agentic flooding w 11 krajach, udział hacktywizmu w incydentach sektora publicznego i luki protokołów A2A — pochodzą z tego korpusu. Pełna lista publikacji znajduje się w manifeście pobrań _NAI_download_manifest.csv biblioteki badawczej TechnoSanacji.

NARZĘDZIA, KTÓRE SŁUŻĄ — NIE SYSTEM, KTÓRY DOMINUJE

TechnoRepublika nie potrzebuje większej liczby ekranów, chatbotów i kont użytkowników. Potrzebuje krótszej drogi od potrzeby do decyzji, od decyzji do wykonania i od błędu do korekty — oraz obywatela, który tę drogę rozumie i może ją skontrolować. Rewolucja agentowa sprawia, że ta droga może być krótsza niż kiedykolwiek. Może też stać się nieprzejrzysta jak nigdy dotąd.

Rozstrzygnięcie nie jest technologiczne. Jest ustrojowe: czy agenci będą narzędziem państwa wobec obywatela, czy narzędziem obu stron wspólnoty politycznej. Symetria narzędzi, sprzężenie przez jawne zobowiązania i mandat człowieka są odpowiedzią TechnoSanacji.

Generacja Z wnosi tempo i narzędzia. Seniorzy wnoszą pamięć i odpowiedzialność. Narzędzia AI mają zamieniać ich wspólny potencjał w wynik dla Rzeczypospolitej.

HASŁO

Agent dla państwa i agent dla obywatela. Jedna brama, jawne reguły, wspólny ślad. Człowiek odpowiada.

Polska, 16.09.2026. — Materiał udostępniony na licencji Creative Commons — Uznanie autorstwa

