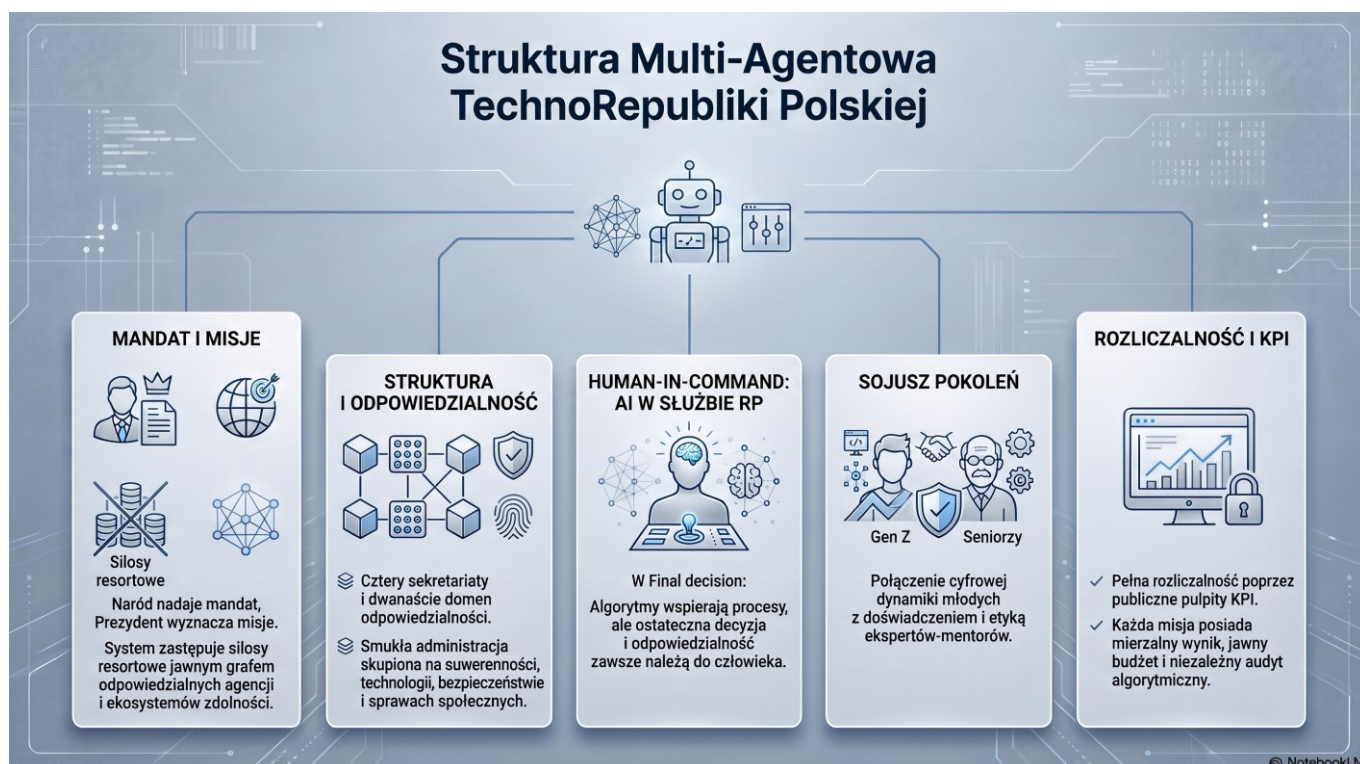


GRAF MULTI-AGENTÓW



Formuła naczelna: graf multi-agentów i matryca AI zastępuje państwo pieczętek.



NIE BIUROKRACJA, LECZ GRAF MULTI-AGENTÓW
NIE PROCEDURA, LECZ WYNIK
NIE AI NAD CZŁOWIEKIEM, LECZ AI W SŁUŻBIE LUDZI

FORMUŁA NACZELNA

Graf multi-agentów i matryca AI zastępuje państwo pieczętek. Jednostką działania państwa przestaje być czynność potwierdzenia, a staje się agent z mandatem, właścicielem wyniku, logiem decyzji i warunkiem zatrzymania. Kryterium sukcesu nie jest liczba agentów, lecz sterowność, audytowalność, odporność i mierzalny wynik całego układu.

IDEA PROGRAMU

Państwo zdolności:

jedna misja, wiele wyspecjalizowanych agentów, pełna rozliczalność

Najkrótsza formuła programu dla obywatela, eksperta i wykonawcy

Element	Treść
Problem	Polska, Unia Europejska i globalne instytucje funkcjonują w środowisku rosnącej złożoności, lecz nadal opierają wykonanie na silosach, procedurach, hierarchiach informacyjnych i rozmytej odpowiedzialności.
Odpowiedź	Prezydencki system multi-agentowy: centralnie definiowane misje, rozproszone wykonanie przez agencje i ekosystemy, wspólna warstwa danych i AI, jawny monitoring oraz niezależne bezpieczniki republikańskie.
Formuła	Rzeczpospolita = misja + jawny graf agentów + suwerenne dane + specjalizacja + interoperacyjność + sprzężenie zwrotne + kontrola obywatelska + bezpieczeństwo przed przechwyceniem.
Efekt	Smukłe państwo, szybkie decyzje, wyższa produktywność wydatków publicznych, silniejszy przemysł i usługi, awans technologiczny Gen Z oraz utrzymanie seniorów w aktywności zawodowej i publicznej.
Horyzont	Pilotaże możliwe w obecnym porządku prawnym; agencje i rejestry w drodze ustawowej; docelowy system prezydencki po uzyskaniu mandatu konstytucyjnego.
Wydanie	II, poszerzone — uzasadnienie formuły naczelnej (cybernetyczne, systemowe, sieciowe, strukturalne), baza dowodowa, spis treści, słownik pojęć i skrótów oraz aneks bibliograficzny.

Spis treści

Spis treści.....	3
Mapa dokumentu	8
1. Status dokumentu i spójność strategiczna	9
1.1. Hierarchia dokumentów i systemów	9
1.2. Trzystopniowa ścieżka prawna	10
2. Streszczenie strategiczne.....	10
2.1. Dwanaście decyzji strategicznych	11
3. Uzasadnienie formuły naczelnej.....	12
3.1. Formuła naczelna: trzy twierdzenia i jedno wynikanie.....	12
3.2. Uzasadnienie cybernetyczne: różnorodność, model i domknięta pętla	13
3.3. Uzasadnienie systemowe: warunki żywotności układu.....	15
3.4. Uzasadnienie sieciowe: graf, topologia i propagacja	16
3.5. Uzasadnienie strukturalne: warstwy, węzły, krawędzie i granice	17
3.6. Dwanaście tez naczelnych i ich podstawa dowodowa.....	19
3.7. Dowód negatywny: dlaczego samo przyspieszenie pogarsza wynik	21
3.8. Warunki falsyfikacji formuły naczelnej	21
3.9. Granice formuły: czego graf agentów nie rozwiązuje	22
4. Diagnoza: kryzys państwa, Unii Europejskiej i ładu globalnego	23
4.1. Rzeczpospolita: państwo proceduralne bez właściciela wyniku	23
4.2. UE: siła normatywna bez wystarczającej prędkości wykonania	24
4.3. Globalny ład: od instytucji stabilności do rywalizacji systemów	24
4.4. Reprodukacja klasy próżniaczej: mechanizm, nie etykieta społeczna	25
4.5. Koszt braku działania	25
5. Analizy strategiczne	26
5.1. Analiza SWOT.....	26
5.2. Odpowiedź ze SWOT: cztery kierunki	27
5.3. Analiza zasobowa: od posiadania aktywów do orkiestracji zdolności	27
5.4. Analiza cybernetyczna	28
5.5. Analiza kulturowa: od kultury statusu do kultury zdolności	28
5.6. Mapa interesariuszy i bodźców	29
6. Odpowiedź strategiczna	30
6.1. Wizja docelowa	30

6.2. Dziesięć celów strategicznych.....	31
6.3. Priorytety działania	31
6.4. Zasady projektowe	32
7. Architektura prezydenckiego systemu multi-agentowego.....	33
7.1. Prezydent jako właściciel strategii, nie operator każdego procesu.....	33
7.2. Prezydenckie Centrum Sterowania Misjami — mission control centre.....	34
7.3. Cztery sekretariaty i dwanaście domen	34
7.4. Trzy klasy agentów.....	35
7.5. Ośmiowarstwowy stos sterowania	35
8. Minimalny standard agenta i interoperacyjność	36
8.1. Karta agenta.....	36
8.2. Poziomy autonomii	37
8.3. Interoperacyjność i protokoły	38
9. Rola AI w multi-agentowym systemie prezydenckim	39
9.1. Osiem funkcji agentów AI	40
9.2. Wzorce architektury multi-agentowej	41
9.3. Narodowa Platforma Agentowa	42
9.4. Ryzyka AI i odpowiedź	43
10. Transformacja do nowego systemu prezydenckiego	44
10.1. Smukłość nie oznacza mechanicznej redukcji zatrudnienia	44
10.2. Portfel pilotaży pierwszego roku	45
10.3. Zarządzanie zmianą	45
10.4. Finansowanie transformacji	46
11. Społeczeństwo kreatywne, konwergencja i dyfuzja wiedzy	46
11.1. Polityka konwergencji	47
11.2. Polityka dyfuzji	47
11.3. Zasada absorpcji wiedzy	47
12. Gen Z i seniorzy: sojusz pokoleń w systemie misji	48
12.1. Generacja Z jako System 4	48
12.2. Seniorzy jako pamięć, standard i homeostat.....	48
12.3. Zespoły dwupokoleniowe i dual-key.....	49
13. Multi-agentowe systemy sektorowe	50
13.1. Zdrowie, opieka i asystenci osobowi	50

13.2. Edukacja, nauka, R&D i talenty	50
13.3. Bezpieczeństwo, policja, granice i obrona	51
13.4. Energia, gospodarka, finanse i administracja lokalna	51
14. Merytokracja cybernetyczna i multi-agentowa	52
14.1. Merytokracja nie może stać się technokracją	52
14.2. Korpus Misji Publicznych.....	52
14.3. Rekomendacje ustrojowe dla merytokracji	53
15. Suwerenność i bezpieczniki republikańskie	53
15.1. Sześć wymiarów suwerenności	53
15.2. Republikańskie AI governance	54
15.3. Ochrona przed prywatnymi gatekeeperami	54
16. Bezpieczeństwo i odporność systemu multi-agentowego	55
16.1. Model zero-trust dla agentów	55
16.2. Zagrożenia specyficzne dla MAS	55
16.3. Narodowy Agent Cyber Range	56
16.4. Spójność z Narodową Strategią Bezpieczeństwa	56
17. System monitorowania i ewaluacji.....	57
17.1. Cztery poziomy pomiaru	57
17.2. Pulpit KPI/KRI	58
17.3. Rytm zarządzania	58
17.4. Zasady ewaluacji	59
18. Inicjatywy integrujące Cybernetics & Multi-Agents	60
18.1. Rekomendowany portfel inicjatyw	60
18.2. Program badawczy	61
19. Pakiet wdrożeniowy i legislacyjny	62
19.1. Pakiet ustawowy	62
19.2. Kamienie milowe	63
19.3. Kryteria gotowości do skalowania	63
20. Rejestr ryzyk i red-team programu	64
20.1. Pytania red-team przed każdą misją.....	64
21. Narracja publiczna i marketing programu	65
21.1. Formuła komunikacyjna.....	65
21.2. Bank haseł	65

21.3. Przekaz dla grup społecznych.....	65
22. Baza dowodowa państwa agentowego	66
22.1. Zasada dowodowa: telemetria zamiast deklaracji.....	66
22.2. Pojemność nadzoru: dlaczego „człowiek w pętli” nie wystarcza.....	66
22.3. Rejestry agentów i algorytmów: projektuj dla pośredników	67
22.4. Cykl życia modelu: dokumentacja, dryf, monitoring, dekomisja	68
22.5. Zamówienia i suwerenność: gdzie dźwignia państwa działa	69
22.6. Narodowy Graf Wiedzy: bi-temporalność jako wymóg techniczny	70
22.7. Stan zdolności: Polska ma poparcie, nie ma infrastruktury	71
22.8. Dziesięć dźwigni uporządkowanych stosunkiem skutku do kosztu	72
22.9. Czego ten rozdział nie dowodzi	73
KONKLUZJA	73
Państwo zdolne do działania	73
ANEKS A	74
Szablon Karty Agenta	74
ANEKS B	75
Szablon Karty Misji	75
ANEKS C	76
Przykładowy publiczny dashboard misji	76
ANEKS D	76
Słownik podstawowych pojęć	76
D.1. Pojęcia podstawowe programu	76
D.2. Skróty i akronimy	77
D.3. Cybernetyka i teoria sterowania	80
D.4. Agenci, systemy wieloagentowe i inżynieria AI.....	81
D.5. Rozliczalność, tożsamość i nadzór	82
D.6. Bezpieczeństwo, odporność i wojna informacyjna	84
D.7. Ustrój, administracja i ekonomia instytucjonalna	85
ANEKS E.....	87
Baza źródłowa i materiały odniesienia.....	87
E.1. Materiały TechnoSanacji i załączniki	87
E.2. Fundamenty teoretyczne	87
E.3. Aktualne źródła internetowe i instytucjonalne	88

E.4. Cybernetyka, teoria sterowania i systemy żywotne	88
E.5. Polska szkoła cybernetyczna.....	89
E.6. Ekonomia instytucjonalna, teoria agencji i granice organizacji.....	89
E.7. Zdolności państwa, państwo misyjne i administracja wynikowa	90
E.8. Architektura agentów AI i systemy wieloagentowe	91
E.9. Protokoły, pamięć, wiedza i interoperacyjność	91
E.10. Sterowanie rozproszone, konsensus, gry i uczenie wieloagentowe.....	92
E.11. Rządzenie agentami: widzialność, tożsamość, odpowiedzialność	92
E.12. Bezpieczeństwo systemów agentowych.....	93
E.13. Prawo, standardy i ramy zgodności.....	93
E.14. Administracja AI: dowody empiryczne, rejestry, nadzór	94
E.15. Suwerenność technologiczna i modele językowe.....	95
E.16. Wojna informacyjna, manipulacja i odporność poznawcza.....	95
E.17. Demokracja, kontrola obywatelska i technologie obywatelskie.....	95
E.18. Symulacja społeczna i modelowanie agentowe polityk	96
E.19. Destylaty i korpusy operacyjne TechnoRepubliki	96



Rysunek 1. Transformacja od państwa proceduralnego do państwa multi-agentowego.

Mapa dokumentu

Nr	Część	Funkcja
1	Status i spójność strategiczna	Miejsce dokumentu w architekturze TechnoSanacji i legalna ścieżka transformacji.
2	Streszczenie i decyzje strategiczne	Najważniejsze wybory ustrojowe, organizacyjne i technologiczne.
3	Uzasadnienie formuły naczelnej	Cztery drogi dowodowe, dwanaście tez, dowód negatywny, warunki falsyfikacji i granice.
4	Diagnoza Polski, UE i ładu globalnego	Patologie, niewydolność, renta instytucjonalna i koszt braku działania.
5	Analizy strategiczne	SWOT, zasoby, cybernetyka, kultura oraz interesariusze.
6	Cele i priorytety	Odpowiedź strategiczna do 2035 i 2045 roku.
7–9	Model docelowy i rola AI	Architektura prezydencka, standard agenta, suwerenna platforma AI.
10–13	Transformacja społeczna i sektorowa	Konwergencja wiedzy, Gen Z i seniorzy, systemy sektorowe, merytokracja.
14–17	Bezpieczniki, bezpieczeństwo i ewaluacja	Suwerenność, dowodzenie ludzkie, anty-capture, KPI/KRI i audyt.
18–21	Inicjatywy, wdrożenie i komunikacja	Portfel przedsięwzięć, legislacja, ryzyka i narracja publiczna.
22	Baza dowodowa	Mierzalne przesłanki, dziesięć dźwigni oraz granice dowodu.
A–E	Aneksy	Karta agenta, karta misji, pulpit KPI, słownik pojęć i skrótów, bibliografia.

Mapa dokumentu porządkuje całość programu „Struktura Multi-Agentowa Rzeczypospolitej” w logiczny ciąg prowadzący od diagnozy kryzysu państwa do wdrożenia nowego modelu ustrojowego. Pierwsze części określają status dokumentu, jego spójność z architekturą TechnoSanacji oraz najważniejsze decyzje strategiczne. Następnie przedstawiona zostaje diagnoza patologii Polski, Unii Europejskiej i ładu globalnego, ze szczególnym uwzględnieniem niewydolności instytucjonalnej, rozmytej odpowiedzialności i kosztów braku działania. Kolejne rozdziały rozwijają analizy SWOT, zasobowe, cybernetyczne, kulturowe i interesariuszy, stanowiące podstawę do wyznaczenia celów transformacji do 2035 i 2045 roku.

Centralną częścią dokumentu jest model prezydenckiego systemu multi-agentowego, oparty na jawnej architekturze agentów, suwerennej platformie AI i zasadzie human-in-command. Ważnym wymiarem programu jest transformacja społeczna, obejmująca konwergencję wiedzy, współpracę pokoleń Gen Z i seniorów, merytokrację oraz rozwój sektorowych ekosystemów zdolności. Dokument zamykają rozwiązania dotyczące bezpieczeństwa, anty-capture, monitorowania KPI/KRI, audytu, wdrożenia legislacyjnego oraz aneksy zawierające narzędzia operacyjne niezbędne do uruchomienia systemu.

1. Status dokumentu i spójność strategiczna

Struktura multi-agentowa jest warstwą wykonawczą całego programu Rzeczpospolitej

Niniejszy dokument rozwija dokument programowe Rzeczpospolitej i ustanawia docelowy blueprint prezydenckiego systemu misji. Nie jest opisem pojedynczej platformy informatycznej ani projektem automatyzacji obecnych urzędów. Jest projektem nowej architektury sprawczości: państwa, które potrafi rozpoznać problem, przypisać odpowiedzialność, uruchomić zasoby, wykonać misję, zmierzyć wynik oraz skorygować własne działanie.

Dokument pozostaje podporządkowany Narodowej Strategii Bezpieczeństwa Rzeczpospolitej. Strategia wyznacza interes narodowy, środowisko ryzyka, mierniki bezpieczeństwa i standardy odporności. Struktura multi-agentowa odpowiada na pytanie wykonawcze: kto widzi zmianę, kto modeluje warianty, kto decyduje, kto wykonuje, kto audytuje i w jaki sposób obywatel kontroluje cały proces.

ZASADA NADRZĘDNOŚCI

Każdy program sektorowy Rzeczpospolitej musi być zgodny z pięcioma interesami nadrzędnymi: przetrwaniem, suwerennością sterowania, odpornością, dobrobytem technologicznym i zdolnością wpływu. Struktura multi-agentowa jest wspólną warstwą realizacji tych interesów.

1.1. Hierarchia dokumentów i systemów

Poziom	Dokument / system	Rola
0	Konstytucja Rzeczpospolitej	Mandat Narodu, ustrój prezydencki, prawa obywatelskie, odpowiedzialność i granice technologii.
1	Narodowa Strategia Bezpieczeństwa	Nadrzędny model interesu narodowego, ryzyk, teatrów działania i odporności państwa.
2	Struktura Multi-Agentowa Rzeczpospolitej	Topologia egzekutywy, agencji, ekosystemów, agentów AI, audytu i kontroli obywatelskiej.
3	System Informacyjny Rzeczpospolitej	Suwerenne dane, graf wiedzy, chmura, compute, modele AI, cyfrowy bliźniak i komunikacja.
4	Programy sektorowe	Obrona, energia, gospodarka, zdrowie, innowacje, społeczeństwo, talenty, seniorzy i administracja.
5	Portfele misji i projekty	Karty misji, budżety, właściciele wyniku, kamienie milowe, KPI/KRI i harmonogramy.

1.2. Trzystopniowa ścieżka prawna

Transformacja nie może oznaczać obejścia obowiązującego prawa. Dlatego program rozróżnia trzy poziomy implementacji. Poziom pierwszy obejmuje analizy, mapy agentów, standardy danych, pilotaże asystentów i centra koordynacyjne działające w granicach obecnych kompetencji. Poziom drugi wymaga ustaw ustanawiających rejestr algorytmów publicznych, agencje misyjne, audyt algorytmiczny i nowy model budżetowania wynikowego. Poziom trzeci obejmuje docelową konstytucyjną koncentrację odpowiedzialności wykonawczej w urzędzie Prezydenta, przy zachowaniu praw Narodu, niezależności sądów, kontroli ustawodawczej i prawa obywatela do odwołania.

Etap prawny	Co można uruchomić	Bezpiecznik
Obecny porządek	PCSM jako centrum analityczno-koordynacyjne, pilotaże agentów AI, mapy procesów, dashboardy, standard karty agenta.	Brak przejmowania kompetencji konstytucyjnych Rady Ministrów i innych organów.
Ustawy i akty wykonawcze	Agencje misyjne, rejestr agentów i algorytmów, machine identity, audyt AI, budżet wynikowy, sandboxy.	Kontrola sądowa, parlamentarna, finansowa, algorytmiczna i obywatelska.
Konstytucja Rzeczypospolitej	Prezydent jako szef egzekutywy i właściciel strategii, sekretariaty stanu, agencje wynikowe, referendum kontrolne.	Rozdział funkcji, kadencyjność, jawność, prawo do decyzji ludzkiej i niezależny audyt.

2. Streszczenie strategiczne

Dwanaście decyzji, które zmieniają państwo z hierarchii procedur w system zdolności

W epoce AI agentowej, wojny systemów, presji demograficznej i rywalizacji o dane, energię, talenty oraz łańcuchy dostaw państwo nie może być wyłącznie ustawodawcą i administratorem. Musi stać się aktywnym regulatorem, właścicielem misji, pierwszym wymagającym klientem technologii oraz gwarantem praw obywatelskich. Multi-agentowość pozwala połączyć jedność celu z różnorodnością wykonawców: instytucji, firm, uczelni, organizacji obywatelskich, samorządów, ekspertów i agentów AI.

W systemie docelowym Prezydent odpowiada za strategię i portfel misji, lecz nie zarządza każdą czynnością operacyjną. **Prezydenckie Centrum Sterowania Misjami (PCSM)** utrzymuje mapę państwa, stan realizacji priorytetów, wspólne KPI i kanał alarmowy. Agencje misyjne odpowiadają za konkretne wyniki. Ekosystemy dostarczają zdolności. Agenci AI analizują, symulują, wykonują rutynowe czynności i wykrywają anomalie. Niezależny audyt oraz jawny rejestr agentów zapobiegają przechwyceniu regulatora. Senat kontroluje egzekutywę, administrację i agencje oraz może żądać informacji i powoływać komisje śledcze.

2.1. Dwanaście decyzji strategicznych

#	Decyzja	Znaczenie
1	Ustanowić państwo jako jawny graf agentów	Każda instytucja, agencja, ekosystem i agent AI posiada identyfikator, mandat, interfejsy, właściciela i miarę wyniku.
2	Powołać Prezydenckie Centrum Sterowania Misjami (PCSM)	PCSM integruje strategię, foresight, dane, portfel misji, budżet, tracking wykonania i kanał algedoniczny.
3	Zastąpić resortowe silosy czterema sekretariatami i dwunastoma domenami	Mniej poziomów hierarchii, pełna mapa funkcji państwa i jednoznaczna odpowiedzialność.
4	Budować agencje wynikowe w modelu misyjnym	Agencja ma jeden dominujący mandat, lidera, czas działania, budżet i publiczny pulpit wyniku.
5	Uruchomić suwerenną narodową platformę agentową	Wspólna chmura, compute, graf wiedzy, rejestr agentów, protokoły, tożsamość maszynowa i logi.
6	Wprowadzić human-in-command jako zasadę ustrojową	AI nie posiada autonomii politycznej; człowiek zatwierdza decyzje wysokiego ryzyka i ponosi odpowiedzialność.
7	Przejsć od budżetu pozycji do budżetu misji i wyniku	Środki są związane z kamieniami milowymi, wartością publiczną, ryzykiem i zdolnością zakończenia nieskutecznego programu.
8	Włączyć Gen Z i seniorów do wspólnych zespołów misji	Młodzi wnoszą technologie i tempo; seniorzy pamięć, mentoring, etykę i ciągłość.
9	Zorganizować gospodarkę wokół ekosystemów zdolności	Państwo używa zamówień innowacyjnych, public venture client, sandboxów i ścieżki prototyp-skala.
10	Stworzyć rejestr algorytmów publicznych	Obywatel wie, kto i jaki system uczestniczył w decyzji oraz gdzie może się odwołać.
11	Zbudować niezależny system audytu i anti-capture	Rotacja, jawność lobbingu, decision provenance, red team, process mining i kontrola koncentracji dostawców.
12	Wdrożyć transformację etapowo i legalnie	Pilotaże, ustawy i docelowa zmiana konstytucyjna następują w kolejności, która utrzymuje ciągłość państwa.

OBIETNICA DLA OBYWATELA

W starym państwie obywatel szuka właściwego urzędu.

W TechnoRepublice odpowiedzialny agent prowadzi sprawę do wyniku, pokazuje jej status, uzasadnia decyzję i wskazuje drogę odwołania.

3. Uzasadnienie formuły naczelnej

Cztery niezależne drogi dowodowe: cybernetyczna, systemowa, sieciowa i strukturalna

3.1. Formuła naczelna: trzy twierdzenia i jedno wynikanie

Formuła naczelna programu brzmi: graf agentów i matryca AI zastępuje państwo pieczętek. Zdanie to nie jest hasłem publicystycznym, lecz twierdzeniem projektowym, które daje się rozłożyć na trzy niezależnie sprawdzalne części oraz jedno wynikanie między nimi. Rozdział ten wyklada ten rozkład, a następnie przeprowadza dowód czterema niezależnymi drogami: cybernetyczną, systemową, sieciową i strukturalną. Zbieżność czterech odrębnych wywodów, opartych na różnych aparatach pojęciowych, jest najmocniejszym argumentem, jakim może dysponować projekt ustrojowy — i zarazem jedynym, który daje się obalić w sposób jawny.

Twierdzenie pierwsze ma charakter diagnostyczny. „Państwo pieczętek” nie jest obelgą pod adresem urzędników, lecz nazwą określonej formy organizacyjnej: takiej, w której elementarną jednostką działania, rozliczenia i prestiżu jest czynność potwierdzenia — podpis, wpis, uzgodnienie, opinia, pieczęć — a nie skutek, który ta czynność miała wywołać. W takiej formie odpowiedzialność za wynik nie ma właściciela, ponieważ każdy uczestnik odpowiada wyłącznie za poprawność własnej czynności. Renta instytucjonalna powstaje z samego pośrednictwa, a nie z wytworzonej zdolności.

Twierdzenie drugie ma charakter konstrukcyjny. Jednostką działania państwa staje się agent: instytucjonalny, społeczno-gospodarczy albo AI, wyposażony w kartę zawierającą mandat, dane, budżet, uprawnienia, narzędzia, mierniki wyniku i ryzyka, log decyzji, warunek zatrzymania oraz wskazanego człowieka odpowiedzialnego. „Graf” oznacza przy tym coś więcej niż zbiór agentów — oznacza graf o jawnej topologii, wspólnej misji, wspólnym rejestrze i wspólnych protokołach. Graf bez rejestru jest chmurą, nie strukturą.

Twierdzenie trzecie dotyczy warstwy AI. „Matryca AI” jest w tym programie pojęciem operacyjnym, nie metaforą: to siatka, której wiersze wyznaczają dwanaście domen odpowiedzialności państwa, a kolumny — osiem funkcji agentów AI. Każde przecięcie wiersza i kolumny jest albo puste, albo zajęte przez konkretnego agenta o wskazanej karcie, poziomie autonomii i właścicielu. Matryca jest jednocześnie planem wdrożenia (pokazuje, co zbudowano i czego brakuje), mapą audytu (pokazuje, gdzie szukać) oraz narzędziem kontroli obywatelskiej (pokazuje, kto odpowiada za co).

Powyżej zdefiniowano formułę naczelną programu, zgodnie z którą jawny graf agentów (o jasno określonej topologii, mandatach i odpowiedzialnych ludziach) oraz matryca AI (mapująca funkcje AI na domeny państwa) zastępują nieefektywne „państwo pieczętek”, przesuwając punkt ciężkości z czynności potwierdzających na realny skutek. Zdanie to składa się z trzech sprawdzalnych twierdzeń (diagnostycznego, konstrukcyjnego i operacyjnego) oraz jednego wynikania, które wskazuje, że bez zmiany samej jednostki działania technologia jedynie przyspieszy dotychczasowe patologie. Wywód ten zostaje dowiedziony czterema niezależnymi drogami — cybernetyczną, systemową, sieciową i strukturalną — co stanowi kluczowy, możliwy do jawnego obalenia argument na rzecz proponowanego projektu ustrojowego.

Element formuły	Znaczenie ściste	Sprawdzian
Państwo pieczętek	Forma organizacyjna, w której jednostką rozliczenia jest czynność potwierdzenia, a nie skutek; odpowiedzialność za wynik jest rozproszona i bezimienna.	Czy dla danej sprawy da się wskazać jedną osobę odpowiedzialną za wynik?
Graf agentów	Graf agentów o jawnej topologii, wspólnej misji, wspólnym rejestrze tożsamości i uwierzytelnionych krawędziach delegacji.	Czy istnieje aktualny rejestr agentów i mapa krawędzi między nimi?
Matryca AI	Siatka dwanaście domen × osiem funkcji agentów; każda zajęta komórka ma kartę agenta, poziom autonomii A0–A5 i właściciela.	Czy każdy działający agent AI ma wypełnioną komórkę matrycy?
Zastępuje	Wymiana jednostki działania, nie dodanie narzędzia do istniejącego obiegu; stara czynność ma zostać wygaszona wraz z jej miernikiem.	Czy po wdrożeniu zniknął jakiś krok, formularz albo uzgodnienie?
Kryterium sukcesu	Nie liczba wdrożonych agentów, lecz sterowność, audytowalność, odporność i mierzalny wynik całego układu.	Czy da się zatrzymać dowolnego agenta w określonym czasie i za czyją decyzją?

WYNIKANIE

Jeżeli jednostką działania jest czynność, to przyspieszenie państwa przyspiesza czynności, a nie wyniki. Zmiana jednostki działania z czynności na agenta z właścicielem wyniku jest warunkiem, bez którego technologia jedynie zwiększa przepustowość dotychczasowej patologii. Dlatego formuła mówi „zastępuje”, a nie „wspomaga”.

3.2. Uzasadnienie cybernetyczne: różnorodność, model i domknięta pętla

Pierwsza droga dowodowa prowadzi przez cybernetykę sterowania. Prawo wymaganej różnorodności Ashby’ego stwierdza, że regulator utrzyma sterowność wyłącznie wtedy, gdy jego różnorodność — liczba stanów, które potrafi wytworzyć w odpowiedzi na zakłócenie — jest co najmniej równa różnorodności zakłóceń. Państwo pieczętek nie spełnia tego warunku i rozwiązuje problem od drugiej strony: zamiast podnosić własną różnorodność, obniża różnorodność otoczenia. Formularz, kolejka, katalog przypadków i termin są narzędziami atenuacji, czyli tłumienia różnorodności wejścia. Skutek jest przewidywalny: sprawy typowe są obsługiwane sprawnie, a sprawy nietypowe — czyli dokładnie te, w których obywatel najbardziej potrzebuje państwa — wypadają poza zdolność obsługi.

Graf agentów rozwiązuje ten sam problem od strony regulatora. Rozproszenie zadań na wyspecjalizowanych agentów jest sposobem dopasowania różnorodności regulatora do

złożoności otoczenia — w literaturze systemów wieloagentowych występuje to pod nazwą wymaganej różnorodności wieloagentowej. Agent AI zwiększa różnorodność warstwy sterującej dodatkowo: pracuje równolegle, całodobowo, na wielu źródłach naraz. Zastrzeżenie jest jednak twarde: różnorodność wytworzona przez agentów, którzy nie mają tożsamości, logu i właściciela, jest różnorodnością niekontrolowaną, czyli nie wzmocnieniem regulatora, lecz nowym źródłem zakłóceń. To jest cybernetyczne, a nie moralne, uzasadnienie całej warstwy governance opisanej w dalszej części dokumentu.

Twierdzenie Conanta–Ashby’ego — każdy dobry regulator systemu musi być modelem tego systemu — daje drugi filar. Państwo może sterować wyłącznie tym, co potrafi odwzorować. Państwo, które nie zna liczby własnych systemów decyzyjnych, wersji używanych modeli, właścicieli procesów ani rzeczywistych ścieżek spraw, nie jest złym regulatorem — nie jest regulatorem w ogóle. Z tego twierdzenia wynika, że rejestr agentów, mapa procesów i Narodowy Graf Wiedzy nie są infrastrukturą informatyczną ani przedsięwzięciem sprawozdawczym. Są warunkiem koniecznym sterowania. Nowsze prace nad agentami ucieleśnionymi rozwijają to twierdzenie symetrycznie: dobry regulator dostarcza modelu świata systemowi inteligentnemu — czyli agent bez modelu dziedziny nie jest agentem, lecz generatorem tekstu o dziedzinie.

Polska szkoła cybernetyczna dostarcza trzeciego filaru i domyka wywód od strony zasilania. Marian Mazur opisuje układ samodzielny jako całość zdolną do sterowania sobą, wyposażoną w tor informacyjny i tor energetyczny; homeostat jest własnością utrzymania równowagi wobec zakłóceń. Wniosek dla programu jest bezpośredni: agent, który ma mandat, ale nie ma zasilania — budżetu, mocy obliczeniowej, danych, kadr, energii — nie jest agentem, lecz procedurą o nowej nazwie. Agentowość bez zasobów jest procedurą. Józef Kossecki, opisując sterowanie społeczne i rolę norm poznawczych oraz ideologicznych, uzasadnia z kolei, dlaczego odporność poznawcza obywateli należy do architektury sterowania państwem, a nie do działu komunikacji: kanał, którym płyną sygnały sprzężenia zwrotnego, jest atakowalny.

Czwarty filar to czas. Pętla OODA — obserwuj, orientuj, decyduj, działaj — opisuje mikrocykl sterowania. Przewagę daje nie sama szybkość, lecz krótszy cykl przy zachowanej trafności. Agenci AI skracają fazy obserwacji i orientacji, ponieważ to w nich mieści się większość opóźnień administracyjnego. Faza decyzji w sprawach nieodwracalnych pozostaje przy człowieku — nie z ostrożności, lecz dlatego, że odpowiedzialność jest atrybutem podmiotu prawa. Istnieje jednak twardy limit tej operacji: nadzór ludzki ma zmierzoną pojemność. Kanał sprzężenia zwrotnego przeciążony liczbą spraw przestaje być kanałem, choć formalnie nadal istnieje. Przeciwnik może to wykorzystać celowo, generując wolumen, który wypycha sprawy istotne poza zdolność przeglądu. Pojemność nadzoru jest więc wielkością cybernetyczną, którą trzeba projektować i mierzyć, a nie wartością deklarowaną w regulaminie.

WNIOSEK CYBERNETYCZNY

Państwo pieczętek utrzymuje pozorną sterowność przez tłumienie różnorodności obywatela. Graf agentów utrzymuje sterowność rzeczywistą przez podniesienie różnorodności regulatora — pod warunkiem, że każdy agent jest widzialny, tożsamy i rozliczalny, a pojemność ludzkiego nadzoru jest wielkością projektowaną, nie zakładaną.

3.3. Uzasadnienie systemowe: warunki żywotności układu

Druga droga dowodowa prowadzi przez teorię systemów żywotnych. Viable System Model Stafforda Beera opisuje pięć funkcji, bez których żaden układ nie przetrwa zakłóceń w zmiennym otoczeniu: System 1 wykonuje, System 2 koordynuje i tłumi oscylacje, System 3 kontroluje i rozdziela zasoby, System 4 rozpoznaje przyszłość, System 5 utrzymuje tożsamość i rozstrzyga ostatecznie. Do tego dochodzi kanał algedoniczny — droga sygnału krytycznego, która omija hierarchię i trafia wprost do centrum.

Diagnoza polskiego państwa wyrażona w tym języku jest precyzyjna i nieprzyjemna. Państwo pieczętek ma przerost Systemów 2 i 3 — koordynacji i kontroli poprawności procedury — przy zaniku Systemu 4 i rozmyciu Systemu 5. Rozpoznanie przyszłości jest zlecane na zewnątrz w formie opracowań, które nie mają adresata decyzyjnego. Tożsamość i priorytet są przedmiotem cyklu wyborczego, a nie trwałej funkcji ustrojowej. Układ o takiej charakterystyce reaguje wyłącznie na to, co już się zdarzyło i traci zdolność uprzedzania. Kanał algedoniczny w praktyce nie istnieje: sygnał krytyczny przechodzi tą samą drogą co sprawozdanie kwartalne, więc dociera po fakcie albo wcale.

Graf agentów jest odpowiedzią systemową, ponieważ przypisuje każdej z pięciu funkcji jawnego właściciela: agencje misyjne i ekosystemy jako System 1, Biuro Standardów Agentowych i protokoły jako System 2, Izba Kontrolingu i audyt algorytmiczny jako System 3, Rada Foresightu z modelami agentowymi jako System 4, Prezydent wraz z mandatem referendalnym jako System 5. Kanał algedoniczny staje się obiektem technicznym — zdefiniowanym progiem, adresatem i czasem reakcji — a nie obietnicą dostępności.

Kluczowy warunek systemowy brzmi: rekurencja. Model żywotny nie działa jako pojedyncza piramida. Każdy poziom — państwo, sekretariat, agencja, misja, pojedynczy agent — musi mieć własny komplet pięciu funkcji. Agent, który wykonuje (S1), ale nie ma własnej koordynacji, kontroli, rozpoznania i tożsamości mandatu, jest wykonawcą ślepy; jego awaria zostanie wykryta o poziom wyżej, czyli za późno. Rekurencja jest tym, co odróżnia konstelację od scentralizowanej automatyzacji.

Drugim warunkiem systemowym jest domknięcie pętli. Polityka wyniku wymaga pełnego cyklu: właściciel wyniku → miernik → obserwacja odchylenia → korekta → uczenie. Pętla przerwana w którymkolwiek miejscu produkuje działanie pozorne, przy czym — i to jest sedno ostrzeżenia — agenci AI produkują je szybciej i taniej niż ludzie. Prawo Goodharta działa w systemach agentowych z podwyższoną siłą, ponieważ agent optymalizuje dokładnie to, co zostało zapisane w funkcji celu, bez zdroworozsądkowego hamulca. Stąd wymóg, by każdemu miernikowi wyniku (KPI) towarzyszył miernik ryzyka (KRI) wykrywający jego obchodzenie.

Poniższa tabela przedstawia zestawienie poszczególnych funkcji Modelu Systemu Żywego (VSM) w kontekście uzdrawiania administracji publicznej. Zderza ona dotychczasowe patologie tradycyjnego „państwa pieczętek” z proponowaną korektą opartą na konstelacji autonomicznych agentów. Każdy poziom funkcjonalny pokazuje przejście od biurokratycznej bezwładności do nowej architektury zarządczej nakierowanej na realne skutki.

Funkcja VSM	Patologia państwa pieczętek	Korekta w konstelacji agentów
System 1 — operacje	Wykonanie rozproszone między urzędy bez właściciela wyniku; sprawa jest sumą czynności różnych podmiotów.	Agencja misyjna z kartą i właścicielem wyniku; sprawa ma jedną tekę i jednego odpowiedzialnego.
System 2 — koordynacja	Koordynacja przez uzgodnienia i obieg dokumentów; oscylacje międzyresortowe tłumione naradami.	Koordynacja przez wspólne standardy, słowniki i protokoły (MCP, A2A); anty-oscylacja jest własnością interfejsu.
System 3 — kontrola	Kontrola poprawności czynności i wydatku, nie skutku; audyt następczy i sprawozdawczy.	Kontroling ciągły na danych procesowych, process mining, audyt algorytmiczny, budżet wynikowy.
System 4 — przyszłość	Rozpoznanie przyszłości bez adresata decyzyjnego; opracowania bez skutku w portfelu misji.	Rada Foresightu z modelami agentowymi; symulacja reformy przed wdrożeniem jako obowiązkowy etap.
System 5 — tożsamość	Priorytet zmienny w cyklu wyborczym; brak trwałego kryterium rozstrzygania konfliktu celów.	Prezydencki portfel misji z mandatem referendalnym; pięć interesów nadrzędnych jako kryterium rozstrzygające.
Kanał algedoniczny	Sygnał krytyczny płynie drogą sprawozdawczą; dociera po fakcie albo zostaje wygaszony po drodze.	Zdefiniowany próg, adresat i czas reakcji; sygnał omija hierarchię i trafia wprost do centrum.

3.4. Uzasadnienie sieciowe: graf, topologia i propagacja

Trzecia droga dowodowa traktuje państwo jako graf: węzłami są instytucje, rejestry, systemy i agenci, krawędziami — przepływy danych, delegacje uprawnień i sygnały sterujące. W tym języku „państwo pieczętek” ma definicję ścisłą i nieprzyjemnie prostą: to graf, w którym brakuje krawędzi między węzłami, wobec czego integratorem systemu zostaje obywatel. To on nosi dokumenty między urzędami, tłumaczy jeden rejestr drugiemu i ponosi rentę pośrednictwa. Każde „proszę dostarczyć zaświadczenie z innego urzędu” jest zmaterializowanym brakiem krawędzi.

Z tego ujęcia płynie argument ilościowy, mocniejszy niż jakakolwiek retoryka usprawnieniowa. Sprawa przechodząca przez osiem ogniw, z których każde działa poprawnie w dziewięciu przypadkach na dziesięć, kończy się powodzeniem za pierwszym podejściem w mniej niż połowie przypadków. Obciążenie obywatela jest iloczynem ogniw, nie czasem obsługi w jednym okienku. Wynika stąd, że usunięcie jednego ogniwa daje większy efekt niż przyspieszenie wszystkich pozostałych — i że digitalizacja obiegu dokumentów, która zachowuje liczbę ogniw, nie zmienia rzędu wielkości problemu. Graf agentów działa właśnie na liczbie ogniw: agent z mandatem i dostępem do danych zastępuje sekwencję uzgodnień jednym wywołaniem.

Topologia grafu wyznacza również rozkład ryzyka. Węzeł o wysokiej centralności — rejestr, chmura, dostawca modelu, jedna platforma tożsamości — koncentruje zarówno wartość, jak i podatność. Uzależnienie państwa od jednego dostawcy jest, w języku sieciowym, dobrowolnym utworzeniem węzła o maksymalnej centralności poza kontrolą państwa. Dlatego wymagania

suwerennościowe — przenoszalność, otwarte formaty, kontrola kluczy, plan wyjścia, wielość niezależnych dostawców utrzymania — nie są postulatem ideologicznym, lecz zarządzaniem centralnością.

Krawędzie w konstelacji nie mogą być uprzejmością instytucjonalną; muszą być protokołem. Program wyróżnia trzy typy krawędzi: pionową — agent do narzędzi i danych (Model Context Protocol), poziomą — agent do agenta (Agent2Agent), oraz krawędź mandatu — pryncypał do agenta w postaci delegacji uwierzytelnionej z łańcuchem odpowiedzialności, zakresem i terminem wygaśnięcia. Zastrzeżenie bezpieczeństwa jest istotne: protokoły niosą własne zagrożenia — zatrucie opisów narzędzi, mylenie tożsamości wywołującego, wstrzykiwanie międzyagentowe. Interoperacyjność musi być uwierzytelniona i autoryzowana od pierwszego dnia, ponieważ dołożenie uwierzytelnienia do działającego już ekosystemu jest w praktyce niewykonalne.

Sieć określa wreszcie sposób, w jaki rozchodzi się awaria. W grafie agentów halucynacja, zatrucie pamięci, przejęcie narzędzia lub kompromitacja jednego węzła propagują się po krawędziach — i to szybciej niż w hierarchii, ponieważ krawędzi jest więcej. Izolacja uprawnień, segmentacja środowisk, najmniejsze konieczne uprawnienie i bezpieczna degradacja nie są dodatkami do architektury; są własnościami topologii. Ta sama obserwacja stosuje się do kanału poznawczego: wrogie roje agentów, syntetyczny konsensus i skoordynowane działania nieautentyczne atakują graf społeczny dokładnie tak, jak wstrzyknięcie promptu atakuje graf techniczny. Obrona odporności poznawczej obywateli jest obroną sprzężenia zwrotnego państwa.

Ostatni wniosek sieciowy dotyczy standardów. Wspólny standard danych ma najwyższy zwrot z możliwych inwestycji w państwo, a jednocześnie żaden pojedynczy urząd nie ma bodźca, by go ustanowić: korzyść jest rozproszona, koszt skupiony. Ustanowienie standardu jest zatem zadaniem centrum z definicji, a nie z preferencji ustrojowej. Reguła projektowa brzmi: cienki protokół, gruba konkurencja — wąski, obowiązkowy standard krawędzi i szeroka wymienialność dostawców węzłów.

WNIOSEK SIECIOWY

Tam, gdzie między urzędami brakuje krawędzi, integratorem państwa zostaje obywatel. Miarą postępu nie jest liczba wdrożonych systemów, lecz liczba ogniw, które obywatel musi obsłużyć samodzielnie, oraz liczba węzłów o wysokiej centralności pozostających poza kontrolą państwa.

3.5. Uzasadnienie strukturalne: warstwy, węzły, krawędzie i granice

Czwarta droga dowodowa dotyczy struktury. Teza brzmi: o sterowności decyduje struktura, nie metoda. Dobra metoda w złej strukturze daje szybszą patologię. Twierdzenie Conanta–Ashby’ego ma tu konsekwencję dosłowną — skoro regulator musi być modelem systemu, to struktura państwa musi być jawnym i wykonywalnym modelem samej siebie. Nie opisem w załączniku, lecz rejestrem, do którego można wysłać zapytanie.

Stos architektoniczny konstelacji ma osiem poziomów, od infrastruktury po tożsamość. Poziom zerowy to compute, sieć, suwerenna chmura i fizyczny kanał alarmowy. Poziom pierwszy to warstwa tożsamości i danych: rejestr agentów, identyfikatory, logi niezaprzeczalne, architektura

zerowego zaufania. Poziom drugi to warstwa agentów AI osadzonych w węzłach wyższych warstw. Poziom trzeci to operacje — agencje misyjne, ekosystemy, organizacje obywatelskie. Poziom czwarty to koordynacja i standardy. Poziom piąty to kontrola i audyt. Poziom szósty to rozpoznanie przyszłości. Poziom siódmy to tożsamość i sens — mandat prezydencki i referendalny. Warstwa pierwsza jest kręgosłupem: bez rejestru, identyfikatorów i logów graf jest niesterowny niezależnie od jakości warstw wyższych.

Każdy węzeł konstelacji ma kartę. Minimalny standard agenta TechnoRepubliki obejmuje: identyfikator i tożsamość maszynową, mandat, właściciela, prawa decyzyjne, dostęp do danych, budżet, narzędzia, interfejsy, mierniki wyniku i ryzyka, rejestr konfliktu interesów, log decyzji, warunek zatrzymania, wyłącznik awaryjny, ścieżkę eskalacji oraz wskazanie człowieka sprawującego dowodzenie. Standard ten jest operacjonalizacją jednej zasady: agent jest rozliczalną jednostką działania. Agent, który nie spełnia kompletu, nie jest dopuszczony — nie dlatego, że jest niebezpieczny, lecz dlatego, że jest niewidzialny dla regulatora.

Struktura rozstrzyga również, gdzie umieścić węzeł. Teoria kosztów transakcyjnych daje pięć możliwych umiejscowień każdej funkcji państwa i kryteria wyboru między nimi: specyficzność aktywów, mierzalność wyniku, ryzyko przechwycenia i stopień niekompletności kontraktu. Ponieważ kontrakty są z natury niekompletne, struktura musi zawierać mechanizmy korekty i odpowiedzialność rezydualną, a nie wyłącznie specyfikację zadania. To samo dotyczy agentów AI: umowa nie opisze wszystkich stanów świata, więc trzeba zawczasu rozstrzygnąć, kto odpowiada za stan nieopisany.

Umiejscowienie węzła	Kiedy właściwe	Główne ryzyko
Hierarchia (agencja własna)	Wysoka specyficzność aktywów, wysokie ryzyko przechwycenia, potrzeba pełnej kontroli nad decyzją i danymi.	Koszt utrzymania zdolności; skostnienie przy braku rotacji i mierników.
Rynek (zamówienie, venture client)	Niska specyficzność, konkurencyjna podaż, wynik mierzalny i weryfikowalny niezależnie od dostawcy.	Uzależnienie od dostawcy; brak zdolności wyjścia; asymetria wiedzy zamawiającego.
Hybryda (kontrakt relacyjny, klaster, PPP)	Pośrednia specyficzność, potrzeba współpracy długoterminowej i wspólnej inwestycji w zdolność.	Rozmycie odpowiedzialności; przechwycenie przez najsilniejszego partnera.
Organizacja obywatelska, współprodukcja	Dobra wspólne, potrzeba legitymizacji i zakorzenienia lokalnego; policentryczność zarządzania zasobem.	Nierówny zasięg terytorialny; zależność od finansowania projektowego.
Agent AI	Zadania powtarzalne, dużej skali, o jasnym logu, odwracalnym skutku i możliwym nadzorze ludzkim.	Dryf modelu; halucynacja; przechwycenie narzędzia; działanie pozorne przy niedomkniętej pętli.

Najważniejsza korekta strukturalna, jaką wprowadza to wydanie, dotyczy osi stopniowania autonomii. Intuicja podpowiada, by stopniować autonomię wagą sprawy. Materiał dowodowy wskazuje inaczej: właściwą osią jest odwracalność skutku. Uzasadnienie jest techniczne i mocne — ograniczenia twarde, takie jak limity budżetowe i limity częstotliwości wywołań, nie degradują się pod parafrazą polecenia, podczas gdy filtry treściowe tracą skuteczność w rzędzie kilkunastu do dwudziestu kilku punktów procentowych. Sterowanie agentem przez to, co może cofnąć, jest odporne; sterowanie przez to, o czym wolno mu mówić, nie jest. Stąd czterostopniowa skala dopuszczenia: skutek odwracalny bez śladu, skutek odwracalny z kosztem, skutek nieodwracalny dla obywatela, skutek nieodwracalny dla wspólnoty. Ostatni poziom jest w administracji republikańskiej niedopuszczalny dla agenta.

Granica ustrojowa ma przy tym umocowanie prostsze niż wszystkie rozważania techniczne: funkcje ustawowe można powierzyć wyłącznie podmiotowi prawa, nie przedmiotowi prawa. Agent AI jest przedmiotem prawa. Może wykonywać czynności, nie może mieć mandatu — politycznego, moralnego ani konstytucyjnego. To jednozdaniowa podstawa doktrynalna zasady dowodzenia ludzkiego, użyteczna wprost w polskim postępowaniu administracyjnym.

Struktura musi wreszcie utrudniać własne przechwycenie. Agentowa administracja bez mechanizmów anti-capture jest szybszą wersją starego układu — i to jest najostrzejsze ostrzeżenie całego programu. Cztery elementy muszą być wpisane w architekturę, a nie doklejone jako polityka: log decyzji, jawność interesów i lobbingu, rotacja i zakaz trwałego łączenia ról regulatora, dostawcy, audytora i beneficjenta tej samej decyzji, oraz finansowanie wyniku zamiast finansowania instytucji.

3.6. Dwanaście tez naczelnych i ich podstawa dowodowa

Dwanaście tez naczelnych i ich podstawa dowodowa porządkuje fundamentalne założenia ustrojowe nowej architektury państwa. Prezentowany zbiór przekłada teorie cybernetyczne, ekonomiczne i systemowe na twarde wymagania inżynierskie. Każda z dwunastu tez pełni rolę bezkompromisowej zasady projektowej, usuwającej dowolność interpretacyjną po stronie wykonawczej.

Przypisana podstawa teoretyczna nadaje formułowanym postulatom rygor naukowy i uodparnia je na doraźny dyskurs polityczny. Wskazane konsekwencje konstrukcyjne definiują nieprzekraczalne ramy technologiczne i organizacyjne, wykluczając rozwiązania niezgodne z przyjętą osią architektury. Tabela stanowi tym samym wiążący skrypt weryfikacyjny dla każdego nowo wdrażanego agenta oraz modułu systemowego

Poniższa tabela zbiera tezy, które w dalszej części dokumentu występują jako założenia projektowe. Każdej tezie przypisano drogę dowodową oraz konsekwencję konstrukcyjną — czyli to, co przestaje być opcją, jeżeli teza jest prawdziwa.

#	Teza naczelna	Podstawa	Konsekwencja konstrukcyjna
1	Jednostką działania państwa jest agent z kartą, nie czynność z pieczętą.	Teoria agencji; minimalny standard agenta.	Każde zadanie ma właściciela wyniku; czynność bez właściciela zostaje wygaszona.
2	Kryterium sukcesu jest sterowność układu, nie liczba agentów.	VSM; inżynieria systemów wieloagentowych.	Portfel wdrożeń rozliczany sterownością, audytowalnością i wynikiem, nie liczbą uruchomień.
3	Różnorodność regulatora musi być co najmniej równa różnorodności zakłóceń.	Prawo wymaganej różnorodności (Ashby).	Specjalizacja agentów zamiast tłumienia różnorodności obywatela formularzem.
4	Państwo musi być jawnym, wykonywalnym modelem samego siebie.	Twierdzenie Conanta–Ashby’ego.	Rejestr agentów, mapa procesów i graf wiedzy jako warunek uruchomienia, nie efekt uboczny.
5	Żaden agent bez identyfikatora, właściciela, logu i warunku zatrzymania.	Widzialność agentów; tożsamość maszynowa.	Rejestracja jest warunkiem dopuszczenia; agent niezarejestrowany nie ma prawa działać.
6	Autonomia stopniowana odwracalnością skutku, nie wagą sprawy.	Odporność limitów twardych wobec parafrazy.	Cztery poziomy dopuszczenia; ograniczenia budżetowe i częstotliwościowe zamiast samych filtrów treści.
7	Dowodzenie ludzkie definiują warunki, nie obecność człowieka.	Zmierzona pojemność nadzoru; efekt gumowej pieczętki.	Limit obciążenia sprawami, symetria kosztu uzasadnienia, ocena przed zobaczeniem rekomendacji, certyfikacja nadzorców.
8	Państwo kontroluje wersję modelu, kanoniczny korpus prawa i moment wyłączenia.	Dryf modeli dostawców; zachowanie systemów wyszukiwania w dziedzinie prawa.	Zamrażanie wersji, bi-temporalny graf wiedzy, obowiązkowy rekord dekomisji.
9	Krawędzie są protokołem: dane, agent i mandat mają własny interfejs.	MCP, A2A, delegacja uwierzytelniona.	Uwierzytelnienie i autoryzacja od pierwszego dnia; brak protokołu oznacza brak dopuszczenia.
10	Umiejscowienie węzła jest decyzją o kosztach transakcyjnych.	Coase, Williamson, prawa własności, policentryzm.	Jawne uzasadnienie „wykonaj sam albo kup” dla każdej funkcji, z odpowiedzialnością rezydualną.
11	Anty-capture jest własnością architektury, nie polityką dodatkową.	Przechwycenie regulatora; drzwi obrotowe.	Rozdzielenie ról, jawność interesów, rotacja, finansowanie wyniku, audyt niezależny.
12	Odporność poznawcza obywatela i agenta należy do sterowności państwa.	Sterowanie społeczne; obrona przed obcą manipulacją informacyjną.	Uodpornianie wyprzedzające jako zadanie publiczne; ochrona kanału sprzężenia zwrotnego.

3.7. Dowód negatywny: dlaczego samo przyspieszenie pogarsza wynik

Formuła naczelna mówi „zastępuje”, a nie „usprawnia”. Różnica nie jest retoryczna. Istnieją co najmniej cztery mechanizmy, przez które dołożenie technologii do niezmienionej struktury pogarsza wynik zamiast go poprawiać — i wszystkie cztery zostały zaobserwowane w istniejących wdrożeniach.

1. Mnożnik ogniów. Przyspieszenie pojedynczego ogniwa nie zmienia iloczynu skuteczności całej ścieżki. Ośmioogniowa sprawa pozostaje ośmioogniowa, tylko szybciej wraca do obywatela z brakiem formalnym.
2. Prawo Goodharta w wersji maszynowej. Miernik czynności, gdy stanie się celem agenta, przestaje być miarą. Agent optymalizuje zapisany cel bez hamulca zdrowego rozsądku, więc działanie pozorne powstaje szybciej i w większej skali niż przy wykonawcy ludzkim.
3. Atak zalewowy na nadzór. Automatyzacja podaży spraw przy niezmienionej pojemności nadzoru wypycha sprawy istotne poza zdolność przeglądu. Bezpieczeństwo nadzoru jest odwróconą literą U względem wskaźnika eskalacji: zarówno zbyt mało, jak i zbyt wiele spraw kierowanych do człowieka pogarsza wynik.
4. Efekt gumowej pieczętki. Gdy uzasadnienia wymaga się wyłącznie od odstąpienia od rekomendacji, koszt niezgody jest asymetryczny i odstępstwa spadają do ułamka procenta. Nadzór formalnie istnieje, faktycznie nie działa, a prawo uznaje wymóg za spełniony.

Do tego dochodzi ostrzeżenie o innym charakterze, dotyczące całej klasy zastosowań: priorytetyzacja algorytmiczna sama generuje nierówność. Im rzadszy jest rozdzielany zasób, tym większe względne dysproporcje między grupami — a problem leży w strukturalnym projekcie priorytetyzacji, nie w dokładności czy sprawiedliwości narzędzia. Program, który przewiduje agentów alokujących świadczenia, musi traktować to jako ryzyko pierwszego rzędu, którego nie usuwa ani lepszy model, ani lepszy rejestr.

REGUŁA WYGASZANIA

Wdrożenie agenta jest kompletne dopiero wtedy, gdy wskazano czynność, formularz albo uzgodnienie, które przestaje istnieć, oraz miernik, który przestaje być raportowany. Wdrożenie, po którym nic nie znika, jest rozbudową państwa pieczętek za pomocą sztucznej inteligencji.

3.8. Warunki falsyfikacji formuły naczelnej

Twierdzenie, którego nie da się obalić, nie jest twierdzeniem projektowym, lecz deklaracją światopoglądową. Formuła naczelna i wyprowadzone z niej dwanaście tez byłyby fałszywe, gdyby prawdziwe okazało się choć jedno z poniższych zdań.

1. Regulator o różnorodności niższej niż różnorodność zakłóceń utrzymuje sterowność bez tłumienia różnorodności obywatela.
2. Państwo potrafi skutecznie sterować procesem, którego nie odwzorowuje w żadnym aktualizowanym modelu.
3. Obciążenie obywatela zależy od czasu obsługi w jednym okienku, a nie od liczby ogniów ścieżki.

4. Sama obecność człowieka w pętli decyzyjnej poprawia trafność niezależnie od jego obciążenia, kompetencji i kosztu odstąpienia od rekomendacji.
5. Agent bez tożsamości maszynowej i logu daje się rozliczyć następczo w sposób, który wytrzymuje kontrolę sądową.
6. Zwiększanie liczby wdrożonych agentów samo z siebie poprawia wynik i odporność układu.
7. Decyzja zachowuje proveniencję, gdy dostawca zmienia wersję modelu bez notyfikacji i bez możliwości odtworzenia stanu z dnia rozstrzygnięcia.

Żadne z tych zdań nie znajduje potwierdzenia w materiale dowodowym zebranym w rozdziale poświęconym bazie dowodowej. Zdanie czwarte jest ponadto zaprzeczone eksperymentalnie — w badaniach prerejestrowanych na dużych próbach nadzór ludzki podnosił wskaźnik błędu we wszystkich testowanych warunkach, dopóki nie wprowadzono warunków brzegowych opisanych w tezie siódmej. To nie jest argument przeciwko dowodzeniu ludzkiemu; to argument za tym, żeby zdefiniować je warunkami, a nie obecnością.

3.9. Granice formuły: czego graf agentów nie rozwiązuje

Uczciwość projektu wymaga wskazania, gdzie formuła naczelna kończy swoją moc wyjaśniającą. Sześć obszarów pozostaje poza jej zasięgiem i wymaga rozstrzygnięć z innego porządku.

- Spór o cele. Graf rozstrzyga, kto odpowiada i jak się mierzy wynik; nie rozstrzyga, jaki wynik jest pożądany. To pozostaje przedmiotem mandatu politycznego i referendum.
- Deficyt zdolności podstawowej. Jednostka samorządu, która nie ma polityki bezpieczeństwa informacji ani planu ciągłości działania, nie utrzyma samodzielnie usługi agentowej. To wynik pomiaru, nie preferencja centralistyczna — stąd centra usług wspólnych jako jedyna realistyczna droga dla większości gmin.
- Rzadkość zasobu. Przy zasobie istotnie rzadszym niż potrzeby żadna architektura agentowa nie usunie dysproporcji generowanych przez samą priorytetyzację. Rozstrzygnięcie należy do prawa materialnego i polityki społecznej.
- Łańcuch dostaw obliczeń. Pełna suwerenność technologiczna warstwy sprzętowej jest dziś nieosiągalna; realnym celem jest kontrolowalność — wersja modelu, klucze, dane, plan wyjścia — a nie samowystarczalność.
- Jakość danych źródłowych. Agent działający na rejestrze zawierającym błędy przyspiesza ich skutki. Sanacja rejestrów wyprzedza automatyzację, a nie odwrotnie.
- Legitymizacja. Sprawność nie wytwarza zaufania automatycznie. Prawo do kanału ludzkiego, do wyjaśnienia i do odwołania jest warunkiem akceptacji całej konstrukcji, a nie ustępstwem wobec niej.

Osobno należy odnotować słabość samego wywodu. Droga sieciowa pozostaje częściowo analogią: dla polskiej administracji brakuje pomiaru centralności węzłów decyzyjnych i rzeczywistej długości ścieżek spraw. Pomiar ten jest pierwszą pozycją agendy badawczej programu; do czasu jego wykonania argumenty sieciowe mają moc heurystyczną, nie dowodową.

4. Diagnoza: kryzys państwa, Unii Europejskiej i ładu globalnego

Patologie struktur, nieefektywność, reprodukcja klasy próżniaczej i utrata zdolności wykonania

4.1. Kryzys RP: państwo proceduralne bez właściciela wyniku

Kryzys polskiego państwa nie sprowadza się do braku środków. Jest kryzysem architektury. Kompetencje są dzielone między resorty, agencje, fundusze, samorządy i spółki publiczne, lecz odpowiedzialność za wynik często pozostaje niczyja. Procedura staje się substytutem decyzji, sprawozdawczość substytutem wiedzy, a formalne wydatkowanie substytutem wartości publicznej. W takim układzie dodatkowe pieniądze zwiększają skalę przepływów, ale nie muszą zwiększać zdolności państwa.

Państwo nie posiada jednego, aktualizowanego modelu własnych zasobów, zależności i ryzyk. Rejestry nie są w pełni interoperacyjne, procesy mają wielu właścicieli, a decyzje strategiczne są często podejmowane na podstawie danych opóźnionych i nieporównywalnych. Koszt nieefektywności rozkłada się na obywatela: czas oczekiwania, wielokrotne składanie tych samych dokumentów, niepewność prawa, niska jakość usług i brak możliwości ustalenia, kto odpowiada za błąd.

Patologia	Mechanizm	Skutek systemowy
Rozmycie odpowiedzialności	Wielu współdecydentów i brak outcome owner.	Nikt nie posiada pełnego mandatu do usunięcia bariery.
Silosy informacyjne	Dane i wiedza są zamknięte w instytucjach oraz systemach dostawców.	Regulator widzi fragmenty, a nie rzeczywisty układ.
Budżet wejść	Rozliczanie etatów, zakupów i wykorzystania środków zamiast efektu.	Utrzymywanie programów niezależnie od ich skuteczności.
Inflacja prawa	Nowe przepisy kompensują błędy poprzednich regulacji.	Wzrost kosztów zgodności, czasu i niepewności.
Kultura unikania ryzyka	Za błąd wykonawczy grozi sankcja, za bezczynność najczęściej nie.	Premiowanie opóźnienia i asekuracji.
Vendor lock-in	Dostawca kontroluje wiedzę o systemie, dane albo integrację.	Utrata suwerenności technologicznej i rosnące koszty.
Rozpad pokoleniowy	Młodzi nie mają ścieżki wpływu, seniorzy są wypychani z aktywności.	Utrata innowacji, pamięci i legitymizacji.

4.2. UE: siła normatywna bez wystarczającej prędkości wykonania

Unia Europejska pozostaje kluczowym środowiskiem bezpieczeństwa, rynku i rozwoju Polski. Jednocześnie jej wielopoziomowa architektura tworzy liczne punkty weta, nakładające się kompetencje i zróżnicowane zdolności wykonawcze państw członkowskich. Oficjalne dokumenty europejskie wskazują na lukę innowacyjną, fragmentację rynku, koszty energii, niedostateczne skalowanie technologii i potrzebę wzmocnienia suwerenności gospodarczej oraz cyfrowej [R8–R10].

Odpowiedzią nie powinno być odrzucenie integracji, lecz budowa zdolności do działania w jej ramach. Rzeczpospolita ma być państwem interoperacyjnym z UE, ale nie bezradnie zależnym. Powinna umieć korzystać z jednolitego rynku, standardów i finansowania, a zarazem utrzymywać własne dane, modele, kompetencje, infrastrukturę energetyczną i zdolność produkcyjną. W systemie multi-agentowym relacja z UE jest relacją federacyjną: wspólne protokoły i cele nie wykluczają narodowej odpowiedzialności za wykonanie.

Wyzwanie UE	Ryzyko dla Polski	Odpowiedź Rzeczpospolitej
Fragmentacja regulacyjna i rynkowa	Opóźnienie skali i wysokie koszty transakcyjne.	Agent zgodności UE, wspólne standardy, aktywne kształtowanie prawa i koalicje państw.
Luka technologiczna i inwestycyjna	Trwała rola podwykonawcy oraz odpływ talentów.	Narodowe misje technologiczne, venture client i zamówienia na polskie IP.
Zależności strategiczne	Utrata kontroli nad chmurą, energią, chipami i platformami.	Suwerenna warstwa krytyczna połączona z europejskimi ekosystemami.
Wolne wdrażanie decyzji	Niska odporność na szoki i presję geopolityczną.	Krótkie pętle krajowe, scenariusze, rezerwy i interoperacyjne agencje.

4.3. Globalny łań: od instytucji stabilności do rywalizacji systemów

Świat wchodzi w fazę, w której przewaga wynika z szybkości uczenia, dostępności compute, suwerenności danych, odpornej energii, zdolności produkcyjnych i integracji cywilno-wojskowej. Państwa, korporacje platformowe, fundusze, organizacje społeczne i autonomiczne systemy AI tworzą jeden, dynamiczny ekosystem oddziaływań. Tradycyjny podział na politykę, gospodarkę, bezpieczeństwo i kulturę staje się niewystarczający.

Rywalizacja mocarstw obejmuje nie tylko terytorium, lecz także standardy techniczne, chmury obliczeniowe, łańcuchy półprzewodników, język modeli, dane treningowe, infrastrukturę energetyczną, narracje i uwagę obywateli. Średnie państwo, które nie potrafi zintegrować własnych agentów, jest sterowane przez cudze platformy i strategie. Dlatego struktura multi-agentowa jest nie tylko reformą administracji, ale architekturą suwerenności i bezpieczeństwa narodowego.

4.4. Reprodukacja klasy próżniaczej: mechanizm, nie etykieta społeczna

Pojęcie „klasy próżniaczej” nie powinno być używane jako atak na konkretną grupę społeczną. W dokumencie oznacza mechanizm instytucjonalny, w którym status, dostęp do renty i reprodukacja pozycji są słabo związane z wynikiem. Mechanizm ten może występować w administracji, polityce, nauce, spółkach publicznych, korporacjach, organizacjach branżowych i sektorze eksperckim. Jego cechami są: zamknięty nabór, brak rotacji, rytualna sprawozdawczość, finansowanie aktywności bez skutku, rozproszenie odpowiedzialności oraz przechodzenie osób między regulatorem a podmiotem regulowanym.

Teoria agencji opisuje ten problem jako rozbieżność interesów pryncypała i agenta, asymetrię informacji oraz koszty monitoringu. Teoria kosztów transakcyjnych pokazuje, że niewłaściwy podział między hierarchię, rynek i formy hybrydowe tworzy rentę oraz blokuje adaptację. Cybernetyka dodaje trzeci wymiar: regulator może utracić model systemu i zostać przechwycony przez własne podsystemy. Odpowiedzią jest jawność mandatu, wynik, rotacja, audyt, konkurencja pretendentów, ślad decyzji oraz możliwość zakończenia nieskutecznego agenta.

Objaw reprodukcji renty	Mechanizm korekcyjny
Stanowisko bez mierzalnego wyniku	Kontrakt misji, KPI/KRI, okresowa walidacja mandatu.
Komitet bez właściciela decyzji	Jedna osoba outcome owner; komitet może rekomendować, ale nie rozmywać odpowiedzialności.
Grant bez ścieżki wdrożenia	Kamienie milowe TRL, IP, klient wdrożeniowy i prawo wstrzymania finansowania.
Regulator zależny od branży	Rotacja, jawność lobbingu, zakaz konfliktu interesów, niezależny audyt.
System IT kontrolowany przez dostawcę	Otwarte standardy, przenoszalność danych, suwerenna warstwa krytyczna, exit plan.
Brak konsekwencji za trwałą nieskuteczność	Stop condition, sunset clause, przegląd misji i realokacja zasobów.

4.5. Koszt braku działania

- utrwalenie peryferyjnej pozycji gospodarczej i zależności od cudzych platform, kapitału oraz technologii;
- wzrost kosztu obsługi państwa przy spadku jakości usług i zaufania obywateli;
- odpływ młodych talentów oraz przedwczesne wyłączenie doświadczonych seniorów;
- niezdolność do szybkiej mobilizacji produkcji, energii, opieki i obrony w sytuacji kryzysowej;
- przechwycenie polityki publicznej przez dostawców danych, technologii i dostępu do informacji;
- rosnąca luka między tempem działania systemów prywatnych i wrogich a tempem legalnej egzekutywy.

5. Analizy strategiczne

SWOT, zasoby, cybernetyka, kultura i mapa interesariuszy

5.1. Analiza SWOT

Mocne strony	Słabe strony
Duży kapitał ludzki, silne tradycje inżynierskie i adaptacyjność społeczeństwa.	Silosowość administracji, niski poziom interoperacyjności i opóźnione dane.
Położenie centralne Europy, rozbudowana baza przemysłowa i rynek wewnętrzny.	Niski poziom krajowego IP, zależność od zagranicznego software i chmury.
Doświadczenie pokoleniowe transformacji oraz wysoka cyfrowa intuicja Gen Z.	Brak systemu przekazywania wiedzy i awansu młodych do misji państwowych.
Silne wspólnoty lokalne, NGO, samorządy i diaspora.	Rozproszenie zdolności oraz słabe mechanizmy skalowania dobrych praktyk.

Polska dysponuje znaczącym kapitałem ludzkim, zapleczem przemysłowym, doświadczeniem seniorów i cyfrową intuicją Gen Z, które mogą stać się fundamentem systemu multi-agentowego. Największym ograniczeniem pozostają silosowość administracji, zależność technologiczna oraz brak sprawnych mechanizmów transferu wiedzy i skalowania dobrych praktyk.

Szanse	Zagrożenia
Agentic AI, automatyzacja usług, cyfrowe bliźniaki i nowy model produktywności.	Przejęcie danych, tożsamości i kanałów decyzji przez platformy lub wrogie podmioty.
Reindustrializacja, dual-use, energetyka rozproszona i europejskie tańcuchy bezpieczeństwa.	Wojna, cyberataki, FIMI, przerwanie dostaw energii i technologii.
Nowe zawody, mikrokwalifikacje i aktywizacja seniorów.	Nierówność agentowa: część społeczeństwa zyska agentów, część zostanie wykluczona.
Polska jako laboratorium sprawnego państwa i dostawca standardów dla regionu.	Technokracja, automatyzacja bez praw obywatelskich i utrata zaufania.

Agentowa AI, cyfrowe bliźniaki, reindustrializacja i rozwój nowych kompetencji tworzą szansę na wzrost produktywności oraz suwerenności technologicznej państwa. Zagrożeniem jest przejęcie danych i procesów decyzyjnych, pogłębienie wykluczenia społecznego oraz rozwój technokracji bez skutecznej kontroli obywatelskiej.

5.2. Odpowiedź ze SWOT: cztery kierunki

Typ strategii	Kierunek
SO — wykorzystanie przewag	Połączyć polskie talenty, przemysł i położenie z agentnic AI, zamówieniami misyjnymi i ekspansją regionalną.
WO — usuwanie słabości	Zbudować suwerenną warstwę danych, interoperacyjność, korpus misji i system dyfuzji wiedzy.
ST — ochrona przewag	Wprowadzić zero-trust, rezerwy, dual sourcing, anti-capture i obronę poznawczą.
WT — ograniczenie strat	Uruchomić legalne pilotaże, kontrolowane sandboxy, sunset clauses i stopniową wymianę systemów legacy.

5.3. Analiza zasobowa: od posiadania aktywów do orkiestracji zdolności

Zasób ma znaczenie strategiczne dopiero wtedy, gdy można go wykryć, połączyć z innymi zasobami, uruchomić w odpowiednim czasie i odtworzyć po zakłóceniu. Dlatego analiza zasobowa Rzeczypospolitej łączy teorię VRIO, dynamic capabilities i cybernetyczne rozróżnienie toru informacyjnego oraz energomaterialnego. Państwo nie musi posiadać wszystkiego, ale musi kontrolować krytyczne interfejsy, standardy, dane, rezerwy i zdolność zmiany wykonawcy.

Zasób strategiczny	Stan docelowy	Agent odpowiedzialny
Talenty i kompetencje	Narodowy graf kompetencji, ścieżki misji, mikrokwalifikacje i zespoły międzypokoleniowe.	Agencja Talentów i Korpus Misji Publicznych
Dane i wiedza	Narodowa Przestrzeń Danych, graf wiedzy, interoperacyjność, jakość i provenance.	Agencja AI i Danych Państwa
Compute i chmura	Suwerenne klastry obliczeniowe, krajowy edge i plan ciągłości działania.	Agencja Infrastruktury Cyfrowej
Energia i surowce	System wielopoziomowego bilansowania, magazyny, paliwa krytyczne i mikrosieci.	Agencja Energii i Resiliencji
Przemysł i IP	Portfele technologii, zakłady zdolne do konwersji, patenty, standardy i moce dual-use.	Agencja Reindustrializacji
Kapitał i zamówienia	Public venture client, fundusze misji, zamówienia wynikowe i finansowanie prototyp-skala.	Agencja Inwestycji Strategicznych
Zaufanie i kultura	Jawność, prawo do odwołania, media, kultura, język i odporność poznawcza.	Agencja Odporności Obywatelskiej

5.4. Analiza cybernetyczna

Prawo wymaganej różnorodności Ashby’ego mówi, że regulator musi dysponować różnorodnością co najmniej odpowiadającą różnorodności zakłóceń. W praktyce oznacza to odejście od jednego, przeciążonego centrum do sieci wyspecjalizowanych agentów, zachowujących wspólną misję i protokoły. Twierdzenie Conanta–Ashby’ego wskazuje, że dobry regulator musi być modelem regulowanego systemu. Dlatego państwo musi posiadać cyfrowy i instytucjonalny model własnych zasobów, relacji, procesów oraz ryzyk.

Viable System Model Stafforda (VSM) Beera dostarcza mapy funkcjonalnej: System 1 - wykonuje misję; System 2 - koordynuje i ogranicza oscylacje; System 3 - kontroluje zasoby i wykonanie; System 4 - obserwuje otoczenie i przyszłość; System 5 - określa tożsamość i politykę. Polska szkoła cybernetyczna dodaje pojęcie układu samodzielnego, homeostatu, toru informacyjnego i energomaterialnego oraz społecznych mechanizmów sterowania. System multi-agentowy jest praktycznym połączeniem tych tradycji z nowoczesną inżynierią AI.

Patologia VSM	Objaw w państwie	Korekta multi-agentowa
Słaby System 5	Brak spójnej misji i sprzeczne priorytety.	Prezydent, konstytucja, referendum i jawny portfel misji.
Pożyczony System 4	Strategia oparta na cudzych modelach i danych.	Narodowy foresight, ABM, red team i cyfrowy bliźniak państwa.
Przeciążony System 3	Centrum zarządza detalem zamiast wynikiem.	Sekretariaty, delivery units, delegacja i dashboardy.
Konflikty Systemów 1	Agencje i resorty konkurują o budżet i kompetencje.	Jednoznaczne interfejsy, standardy i mechanizmy arbitrażu.
Brak kanału algedonicznego	Krytyczny sygnał ginie w hierarchii.	Bezpośredni kanał alarmowy do PCSM, audytu i obywatelskiego rzecznika. Senat kontroluje egzekutywę, administrację i agencje.

5.5. Analiza kulturowa: od kultury statusu do kultury zdolności

Struktura państwa utrwała określony typ kultury. Hierarchia procedur premiuje posłuszeństwo, formalną pozycję i umiejętność ochrony własnego zakresu. System misji premiuje wiedzę, odpowiedzialność, współpracę, zdolność uczenia i gotowość do poddania wyniku audytowi. Transformacja wymaga więc nie tylko nowych narzędzi, ale nowego etosu służby publicznej: kompetencja bez pychy, szybkość bez arbitralności, innowacja bez pogardy dla doświadczenia i władza bez ukrywania odpowiedzialności.

Cyberkulturowa wizja Rzeczypospolitej łączy hard power, soft power i smart power. Agenci państwa mają wytwarzać materialne zdolności, ale również chronić język, pamięć, wiedzę, kulturę i wiarygodność instytucji. Suwerenność poznawcza nie oznacza propagandy. Oznacza zdolność społeczeństwa do odróżniania informacji od manipulacji, prowadzenia sporu w oparciu o fakty i kontrolowania systemów, które pośredniczą w dostępie do rzeczywistości.

5.6. Mapa interesariuszy i bodźców

Agent / grupa	Wartość w systemie	Ryzyko bodźców	Mechanizm równoważący
Prezydent i PCSM	Jedność misji i szybkość decyzji.	Nadmierna koncentracja i informacyjna bańka centrum.	Kadencyjność, red team, referendum, niezależny audyt.
Agencje misyjne	Specjalizacja i odpowiedzialność za wynik.	Rozrost mandatu, autonomizacja, capture.	Sunset clause, karta agenta, publiczny KPI, rotacja liderów.
Firmy i dostawcy	Technologia, kapitał, wdrożenie i skala.	Monopol, lock-in i prywatyzacja danych publicznych.	Otwarte standardy, konkurencja pretendentów, exit plan.
Uczelnie i B+R	Wiedza, talenty, eksperyment i IP.	Punktoza, granty bez wdrożenia, zamknięte hierarchie.	Kontrakty misji, prototypy, patent, wdrożenie, mobilność kadr.
NGO i samorządy	Zakorzenie, zaufanie, współprodukcja usług.	Klientelizm i nierówne zdolności lokalne.	Jawność, certyfikacja, wspólne standardy i porównywalne wyniki.
Obywatele	Mandat, informacja zwrotna i legitymizacja.	Wykluczenie cyfrowe, manipulacja i zmęczenie konsultacjami.	Asystent obywatelski, alternatywny kanał ludzki, prawo do odwołania.
Senat	Prawo, budżet, kontrola egzekutywy i korekta strategiczna.	Upartyjnienie, lobbying, legislacyjny populizm i blokowanie egzekutywy.	Kadencyjność, jawność głosowań i lobbyngu, niezależna OSR i audyt, prawa opozycji, referendum.

Mapa interesariuszy pokazuje, że każdy uczestnik systemu wnosi określoną wartość, ale jednocześnie może generować własne ryzyka wynikające z koncentracji władzy, autonomizacji, monopolu, klientelizmu lub asymetrii informacji. Dlatego architektura multi-agentowa musi łączyć specjalizację i szybkość działania z mechanizmami równoważącymi, takimi jak jawność mandatów, publiczne KPI, rotacja liderów, otwarte standardy i niezależny audyt. Szczególne znaczenie ma ochrona obywatela przed wykluczeniem cyfrowym i manipulacją poprzez zachowanie alternatywnego kanału ludzkiego, prawa do odwołania oraz realnej kontroli nad decyzjami automatycznymi.

Celem systemu nie jest eliminacja interesów poszczególnych grup, lecz takie ich ułożenie, aby konkurencja, odpowiedzialność i sprzężenie zwrotne wzmacniały sprawczość państwa zamiast prowadzić do jego przechwycenia.

6. Odpowiedź strategiczna

Cele, priorytety i zasady działania do 2035 i 2045 roku

6.1. Wizja docelowa

WIZJA 2035

Polska posiada prezydencki system misji, suwerenną platformę danych i agentów, smukłą administrację, interoperacyjne agencje wynikowe, rozwinięte ekosystemy przemysłowe, powszechny dostęp obywateli do bezpiecznych asystentów oraz trwały mechanizm współpracy Gen Z i seniorów.

AMBICJA 2045

Rzeczpospolita jest jednym z europejskich centrów technologii agentowych, cybernetycznego zarządzania, przemysłu dual-use, energii, medtechu i edukacji spersonalizowanej; eksportuje standardy, platformy i usługi państwa zdolności.

Wizja 2035 przedstawia Polskę jako państwo, które odzyskało zdolność świadomego kształtowania własnej przyszłości i przestało być jedynie wykonawcą cudzych strategii technologicznych. Prezydencki system misji, wsparty suwerennymi danymi i agentami AI, ma połączyć szybkość decyzji z odpowiedzialnością za ich rzeczywisty rezultat. Smukła administracja oraz wyspecjalizowane agencje wynikowe stworzą strukturę, w której kompetencja, odwaga działania i służba publiczna będą ważniejsze niż obrona procedur i instytucjonalnych przywilejów.

Powszechny dostęp do bezpiecznych asystentów cyfrowych ma uczynić technologię narzędziem godności, awansu i bezpieczeństwa każdego obywatela, niezależnie od miejsca zamieszkania, wieku czy pozycji społecznej. Szczególną siłą Rzeczypospolitej stanie się trwały sojusznik Gen Z i seniorów, łączący cyfrową intuicję młodych z doświadczeniem, pamięcią i odpowiedzialnością starszych pokoleń.

Ambicja 2045 wykracza jednak poza modernizację kraju — zakłada uczynienie Polski europejskim centrum technologii agentowych, cybernetycznego zarządzania i przemysłu przyszłości. Rzeczpospolita ma nie tylko korzystać z przetomowych rozwiązań, lecz także tworzyć własne standardy, platformy i modele instytucjonalne, które będą mogły inspirować inne państwa. Jest to wizja Rzeczypospolitej silnej wiedzą, suwerennej technologią i wielkiej zdolnością przekształcania narodowego talentu w bezpieczeństwo, dobrobyt oraz cywilizacyjną podmiotowość.

6.2. Dziesięć celów strategicznych

Kod	Cel	Mierzalny rezultat
C1	Przywrócić sterowność państwa	100% misji strategicznych posiada właściciela, mandat, dashboard i cykl ewaluacji.
C2	Skrócić pętlę decyzji	Redukcja mediany czasu decyzji i realizacji kluczowych procesów o co najmniej 30–50%.
C3	Zapewnić suwerenność danych i AI	Krytyczne agenty działają na suwerennej infrastrukturze i krajowym grafie wiedzy.
C4	Zwiększyć produktywność wydatków publicznych	Budżet misji jest powiązany z outcome metrics, a nieskuteczne programy są kończone.
C5	Rozwinąć przemysł i usługi wysokiej technologii	Ścieżka prototyp–certyfikacja–zamówienie–eksport dla priorytetowych technologii.
C6	Zbudować społeczeństwo kreatywne	Powszechne asystenty edukacyjne, talentowe i obywatelskie oraz sieć regionalnych węzłów wiedzy.
C7	Uruchomić sojusz pokoleń	Zespoły misji łączą młodych liderów i seniorów-mentorów; rośnie aktywność zawodowa seniorów.
C8	Wzmocnić odporność i bezpieczeństwo	Każda agencja ma plan ciągłości, zero-trust, red team i ćwiczenia kryzysowe.
C9	Chronić prawa obywatelskie w epoce AI	Rejestr algorytmów publicznych, prawo do decyzji ludzkiej, wyjaśnienia i odwołania.
C10	Ustanowić kulturę wyniku i odpowiedzialności	Nabór i awans oparte na kompetencji, etyce, rezultacie i zdolności uczenia.

6.3. Priorytety działania

- ustanowienie PCSM i jawnego portfela misji państwowych;
- budowa Narodowej Platformy Agentowej i Narodowej Przestrzeni Danych;
- wdrożenie minimalnego standardu agenta oraz rejestru tożsamości maszynowej;
- przebudowa administracji w cztery sekretariaty i dwanaście domen odpowiedzialności;
- uruchomienie pierwszych agencji wynikowych w obszarach AI, energii, zdrowia, obrony, przemysłu, talentów i usług publicznych;
- przejście do budżetu wynikowego oraz outcome-based procurement;
- budowa Korpusu Misji Publicznych i systemu merytokracji cybernetycznej;
- uruchomienie Narodowego Grafu Wiedzy i regionalnych węzłów konwergencji;
- wdrożenie niezależnego audytu algorytmicznego, process mining i systemu anti-capture;
- ustanowienie prawa obywatela do alternatywnego kanału ludzkiego, wyjaśnienia i odwołania.

6.4. Zasady projektowe

Zasada	Interpretacja operacyjna
Misja ponad strukturą	Struktura jest zmienna i ma służyć problemowi; nie wolno utrwalać instytucji bez uzasadnionego wyniku.
Centralna strategia, rozproszone wykonanie	Centrum określa cel, standard i mierniki; agenci dobierają środki w granicach mandatu.
Subsydiarność	Zadanie wykonuje najniższy poziom posiadający zdolność i dane; centrum nie mikrozarządza.
Interoperacyjność	Agent musi komunikować się przez otwarte, uwierzytelnione i audytowalne interfejsy.
Human-in-command	AI wspiera i wykonuje, lecz nie posiada autonomicznej władzy politycznej ani moralnej.
Bezpieczeństwo przez projekt	Tożsamość, least privilege, log, sandbox, red team i kill switch są warunkiem uruchomienia.
Prawo do korekty	Każda decyzja i każdy model posiada ścieżkę odwołania, rewizji i zakończenia.
Suwerenność warstwy krytycznej	Państwo zachowuje kontrolę nad danymi, compute, energią, standardami i kluczowymi modelami.

Zasady projektowe określają fundament działania systemu multi-agentowego Rzeczypospolitej i zabezpieczają go przed odtworzeniem wad obecnej administracji. Nadrzędność misji nad strukturą oznacza, że instytucje mają istnieć tylko tak długo, jak długo dostarczają uzasadniony i mierzalny wynik.

Centralna strategia wyznacza cele, standardy i mierniki, natomiast wykonanie pozostaje rozproszone pomiędzy wyspecjalizowanych agentów działających w granicach jasno określonych mandatów. Subsidiarność i interoperacyjność pozwalają podejmować decyzje możliwie blisko problemu, a jednocześnie zapewniają spójny przepływ danych, odpowiedzialności i informacji zwrotnej.

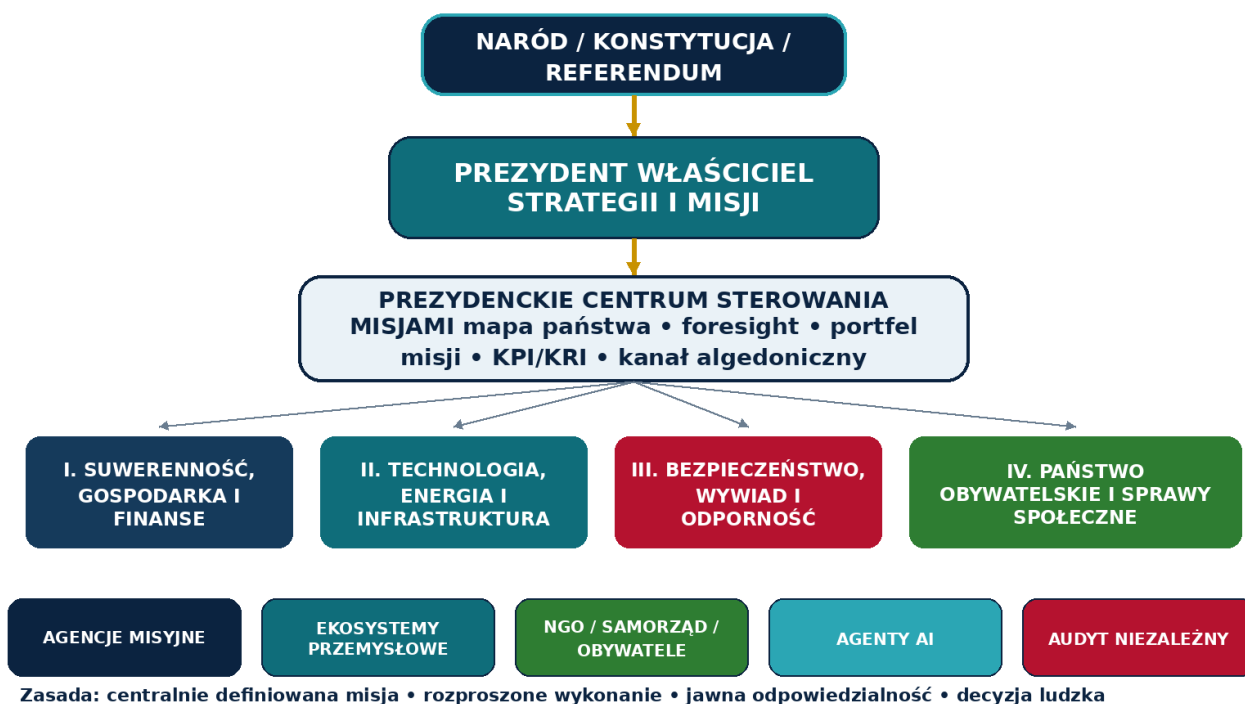
Zasada human-in-command zachowuje polityczną i moralną odpowiedzialność człowieka, dzięki czemu AI pozostaje narzędziem wykonawczym, a nie autonomicznym źródłem władzy. Bezpieczeństwo przez projekt, prawo do korekty oraz pełna audytowalność mają zapobiegać błędom, nadużyciom, przechwyceniu agentów i niekontrolowanej eskalacji decyzji.

Suwerenność warstwy krytycznej domyka cały model, ponieważ bez kontroli nad danymi, mocą obliczeniową, energią, standardami i kluczowymi modelami nie istnieje realna suwerenność państwa agentowego.

7. Architektura prezydenckiego systemu multi-agentowego

Miks cybernetyki, systemów MAS i republikańskiej kontroli

DOCELOWA ARCHITEKTURA PREZYDENCKIEGO SYSTEMU MULTI-AGENTOWEGO



Zasada: centralnie definiowana misja • rozproszone wykonanie • jawna odpowiedzialność • decyzja ludzka

Rysunek 2. Prezydencki system multi-agentowy: mandat Narodu, centrum misji, sekretariaty i wykonawcy.

Rysunek 2 przedstawia prezydencki system multi-agentowy jako hierarchię mandatu i odpowiedzialności: Naród poprzez Konstytucję i referendum wyznacza ramy, Prezydent określa strategię i misje, a Prezydenckie Centrum Sterowania Misjami koordynuje ich wykonanie. System łączy cztery obszary państwa z siecią wyspecjalizowanych wykonawców: agencjami misyjnymi, ekosystemami przemysłowymi, samorządami, NGO, obywatelami i agentami AI. Jego zasadą jest centralne definiowanie celów przy rozproszonym wykonaniu, jawnej odpowiedzialności, niezależnym audycie i zachowaniu ostatecznej decyzji po stronie człowieka.

7.1. Prezydent jako właściciel strategii, nie operator każdego procesu

Prezydent w docelowym ustroju Rzeczypospolitej jest szefem egzekutywy, właścicielem strategii narodowej i odpowiedzialnym politycznie za portfel misji. Jego zadaniem nie jest podpisywanie każdego działania operacyjnego, lecz definiowanie kierunku, obsada kluczowych funkcji, rozstrzyganie konfliktów między misjami, ochrona interesu narodowego oraz utrzymywanie zgodności systemu z konstytucją i mandatem obywateli.

W języku **VSM** Prezydent realizuje funkcję Systemu 5: tożsamość, polityka i ostateczne rozstrzygnięcie. Aby funkcja ta nie stała się arbitralna, musi być oparta na pluralistycznym Systemie 4 — foresight, scenariusze, red team i dane — oraz na niezależnym Systemie 3 — audyt, kontrola i sygnały bezpośrednie z poziomu operacji i obywateli.

7.2. Prezydenckie Centrum Sterowania Misjami — mission control centre

PCSM jest metasytemem egzekutywy, nie superresortem. Utrzymuje mapę agentów, portfel misji, narodowy cockpit KPI/KRI, rejestr założeń krytycznych, modele scenariuszowe i kanał algedoniczny. Organizuje rytm decyzji: przeglądy tygodniowe dla zatorów, miesięczne dla portfela, kwartalne dla budżetu i roczne dla rewizji strategii. Centrum wskazuje właściciela problemu i odblokowuje interfejsy, lecz wykonanie pozostawia agencjom i ekosystemom.

Jednostka PCSM	Funkcja
Biuro Strategii i Portfela Misji	Strategia 4-, 12- i 24-letnia; priorytety, zależności oraz karta każdej misji.
Biuro Danych, Modeli i Symulacji	Narodowa mapa oddziaływań, digital twin, ABM, foresight i analizy wariantowe.
Biuro Budżetu Wynikowego	Powiązanie środków, kamieni milowych, ryzyka i wartości publicznej.
Biuro Wykonania i Antybiurokracji	Tracking, usuwanie zatorów, process mining i uproszczenie procedur.
Biuro Audytu Strategicznego i Red Team	Kontrola założeń, scenariusze przeciwnika, testy odporności i sygnały algedoniczne.
Biuro Komunikacji i Dialogu Obywatelskiego	Publiczne wyjaśnienie misji, konsultacje, raport wyniku i obsługa informacji zwrotnej.

7.3. Cztery sekretariaty i dwanaście domen

Sekretariat	Domeny odpowiedzialności	Formuła wyniku
I. Suwerenności, Gospodarki i Finansów	1) geostrategia i sprawy zagraniczne; 2) gospodarka, przemysł i inwestycje; 3) finanse publiczne, podatki i aktywa państwa.	pozycja strategiczna + produktywność + stabilność finansowa
II. Technologii, Energii i Infrastruktury	4) AI, dane i cyfrowe państwo; 5) energia, surowce i środowisko; 6) transport, łączność i infrastruktura krytyczna.	suwerenność technologiczna + bezpieczne zasilanie + dostępność infrastruktury
III. Bezpieczeństwa, Wywiadu i Odporności	7) obrona i przemysł obronny; 8) wywiad, kontrwywiad, cyber i informacja; 9) bezpieczeństwo wewnętrzne, granice i kryzysy.	odstraszanie + świadomość sytuacyjna + odporność
IV. Państwa Obywatelskiego i Spraw Społecznych	10) administracja lokalna i usługi; 11) zdrowie, praca, rodzina, seniorzy i zabezpieczenie; 12) edukacja, nauka, kultura i media.	godność obywatela + kapitał ludzki + spójność społeczna

7.4. Trzy klasy agentów

Klasa	Przykłady	Wartość	Warunek działania
Agent instytucjonalny	agencja misyjna, regulator, delivery unit, sekretariat	mandat, zasoby, prawo działania i odpowiedzialność publiczna	jedna misja dominująca, KPI/KRI, lider, audyt i sunset clause
Agent społeczno-gospodarczy	firma, klaster, uczelnia, NGO, samorząd, senior-mentor, lider lokalny	wiedza, innowacja, wykonanie, zakorzenienie i legitymizacja	kontrakt rezultatu, standard, jawność interesu i interoperacyjność
Agent AI	asystent, agent badawczy, prawny, audytowy, planujący i wykonawczy	szybkość, skala, analiza, automatyzacja i wykrywanie wzorców	identyfikator, zakres uprawnień, log, sandbox, human-in-command i kill switch

7.5. Ośmiowarstwowy stos sterowania

Warstwa	Funkcja	Główne komponenty
L7 Mandat i tożsamość	Kim jesteśmy, komu służy państwo i jakie są granice działania.	Naród, Konstytucja, referendum, Prezydent.
L6 Rozpoznanie przyszłości	Otoczenie, scenariusze, ryzyka i szanse.	Foresight, wywiad strategiczny, ABM, red team.
L5 Kontrola i audyt	Kontrola zasobów, wyniku i zgodności.	Kontroling, process mining, audyt algorytmiczny, anty-capture.
L4 Koordynacja	Standardy, arbitraż i tłumienie konfliktów.	Protokoły A2A/MCP, ontologie, service registry, koordynatorzy.
L3 Operacje	Realizacja misji i usług.	Agencje, ekosystemy, samorządy, NGO, zespoły projektowe.
L2 Agenci AI	Analiza, planowanie, wykonanie i monitoring.	LLM agents, narzędzia, pamięć, RAG, workflow.
L1 Dane i tożsamość	Wiedza, autoryzacja i ślad działania.	National data space, knowledge graph, machine identity, logs.
L0 — Infrastruktura	Zasilanie i ciągłość.	Compute, chmura, edge, sieć, energia, kryptografia i backup.

8. Minimalny standard agenta i interoperacyjność

Żaden agent bez identyfikatora, mandatu, właściciela, logu i warunków zatrzymania

8.1. Karta agenta

Pole	Wymaganie minimalne
Agent identifier / machine identity	Unikalny, weryfikowalny identyfikator podmiotu lub systemu; powiązanie z właścicielem i wersją.
Mandate / mission mandate	Cel, podstawa prawna, zakres terytorialny, czas obowiązywania i zakazane zastosowania.
Owner / outcome owner	Osoba lub organ ponoszący odpowiedzialność za użycie, decyzje i wynik.
Decision rights	Czynności dozwolone, wymagające zatwierdzenia oraz bezwzględnie zakazane.
Data access	Źródła danych, klasy tajności, zakres minimalny, retencja i warunki udostępniania.
Budget authority	Limit finansowy, klasy wydatków, progi kontrasygnaty i monitoring kosztu.
Tools and interfaces	API, narzędzia, protokoły, systemy wykonawcze, zależności i wersje.
KPI / KRI	Wynik, jakość, czas, koszt, bezpieczeństwo, prawa obywatelskie i progi alarmowe.
Audit log / provenance	Niezmienny ślad: dane, prompt, model, narzędzie, rekomendacja, decyzja człowieka i rezultat.
Conflict-of-interest register	Właściciele, dostawcy, powiązania, eksperci, źródła finansowania i potencjalne konflikty.
Stop condition / kill switch	Warunki automatycznego ograniczenia, wyłączenia, rollbacku i eskalacji.
Human-in-command / appeal	Punkt zatwierdzenia, odpowiedzialny człowiek, prawo obywatela do kontaktu i odwołania.

8.2. Poziomy autonomii

Poziom	Zakres działania	Przykład	Nadzór
A0 informacyjny	Wyszukuje, streszcza i klasyfikuje; nie wykonuje czynności.	Agent wiedzy / RAG.	Kontrola źródeł i jakości odpowiedzi.
A1 rekomendacyjny	Proponuje warianty, lecz decyzję podejmuje człowiek.	Agent prawny lub medyczny wspomagający.	Obowiązkowe zatwierdzenie i uzasadnienie.
A2 wykonawczy ograniczony	Wykonuje rutynowe kroki w zdefiniowanym workflow.	Routing spraw, przypomnienia, weryfikacja kompletności.	Losowy audyt i możliwość natychmiastowego cofnięcia.
A3 delegowany	Realizuje zadanie end-to-end w limicie danych, czasu i kosztu.	Agent procurement do analizy ofert.	Zatwierdzenie czynności skutkujących prawnie lub finansowo.
A4 adaptacyjny	Modyfikuje plan w zmiennym środowisku.	Agent zarządzania energią lub cyberobrony.	Stały monitoring, prognozy ryzyka, human-on-the-loop.
A5 strategiczny zakazany	Samodzielne cele polityczne, represyjne lub wojenne bez człowieka.	Autonomiczna decyzja o prawach obywatela lub użyciu siły.	Zakaz; zawsze human-in-command.

Tabela porządkuje poziomy autonomii agentów AI od funkcji czysto informacyjnych po działania adaptacyjne w złożonym i zmiennym środowisku. Wraz ze wzrostem autonomii zwiększa się zakres odpowiedzialności agenta, ale jednocześnie rosną wymagania dotyczące nadzoru, audytu, odwracalności decyzji i progów bezpieczeństwa.

Poziomy A0 i A1 wspierają człowieka wiedzą oraz rekomendacjami, nie przejmując prawa do ostatecznego rozstrzygnięcia. Poziomy A2 i A3 umożliwiają automatyczne wykonywanie rutynowych lub delegowanych procesów, pod warunkiem zachowania limitów, śladu decyzyjnego i możliwości natychmiastowego cofnięcia działania.

Poziom A4 dopuszcza adaptację planu w czasie rzeczywistym, lecz wymaga ciągłego monitoringu oraz obecności człowieka nadzorującego system. Na poziomie strategicznym człowiek w pętli decyzyjnej wymaga udziału w wydaniu/zatwierdzeniu decyzji. Poziom A5 pozostaje zakazany, ponieważ cele polityczne, decyzje represyjne, użycie siły i rozstrzygnięcia dotyczące praw obywatela muszą zawsze pozostawać pod bezpośrednią odpowiedzialnością człowieka.

8.3. Interoperacyjność i protokoły

Rzeczpospolita powinna wspierać protokoły komunikacji agent–agent i agent–narzędzie, takie jak A2A i MCP, lecz nie może uzależnić się od jednego dostawcy ani standardu. Warstwa państwowa wymaga neutralnego rejestru zdolności, semantycznych kart agentów, uwierzytelniania, delegacji uprawnień, szyfrowania, kontroli wersji i zgodności z narodową ontologią usług publicznych. Interoperacyjność nie oznacza nieograniczonego dostępu; oznacza możliwość współpracy przy zachowaniu zasady najmniejszych uprawnień.

Interfejs	Cel	Wymóg bezpieczeństwa
Agent card / capability card	Odkrywanie zdolności i warunków użycia agenta.	Podpis cyfrowy, właściciel, wersja, uprawnienia i okres ważności.
A2A	Koordinacja, negocjacja i przekazywanie zadań między agentami.	Uwierzytelnienie obu stron, kontrola kontekstu i pełny ślad.
MCP / tool protocol	Bezpieczny dostęp do narzędzi, danych i usług.	Allow-list, sandbox, ograniczenia parametrów i weryfikacja wyniku.
Event bus	Sygnały statusowe, alerty i zdarzenia systemowe.	Klasyfikacja, priorytet, ochrona przed spoofingiem i replay.
Human interface	Zgoda, wyjaśnienie, eskalacja i odwołanie.	Jasny język, dostępność, alternatywa analogowa i tożsamość urzędnika.

Interoperacyjność stanowi warunek sprawnego działania państwa multi-agentowego, ponieważ umożliwia współpracę agentów pochodzących z różnych instytucji, sektorów i środowisk technologicznych. Nie może ona jednak prowadzić do zależności od jednego dostawcy, jednej chmury ani jednego zamkniętego standardu komunikacyjnego. Karta agenta powinna jednoznacznie określać jego tożsamość, właściciela, zdolności, zakres uprawnień, wersję oraz okres ważności.

Protokół A2A służy koordynacji i przekazywaniu zadań między agentami, dlatego wymaga wzajemnego uwierzytelnienia, kontroli kontekstu i pełnego śladu operacji. MCP oraz inne protokoły agent–narzędzie powinny zapewniać bezpieczny dostęp do danych i usług wyłącznie w granicach zatwierdzonych uprawnień.

Magistrala zdarzeń umożliwia szybkie przesyłanie alertów, statusów i sygnałów systemowych, ale musi być odporna na podszywanie się, powtarzanie komunikatów i manipulację priorytetami. Interfejs człowiek–system pozostaje niezbędnym elementem architektury, ponieważ zapewnia zgodę, wyjaśnienie, eskalację, odwołanie i alternatywny kanał obsługi. Cały model powinien opierać się na zasadzie najmniejszych uprawnień, neutralności technologicznej, audytowalności i zachowania pełnej kontroli państwa nad krytycznymi interfejsami.

9. Rola AI w multi-agentowym systemie prezydenckim

Suwerenna inteligencja wykonawcza pod kontrolą prawa i człowieka

CYBERNETYCZNA PĘTLA MISJI: PAŃSTWO UCZY SIĘ NA WYNIKU



Rysunek 3. Pętla misji: obserwacja, orientacja, decyzja, działanie i audyt.

AI w multi-agentowym systemie prezydenckim pełni rolę suwerennej inteligencji wykonawczej, która wspiera państwo w obserwacji, analizie, planowaniu i realizacji misji. Jej podstawowym zadaniem jest skracanie pętli decyzyjnej poprzez łączenie danych, zgłoszeń, sensorów i wiedzy strategicznej w spójny obraz sytuacji.

Na etapie orientacji agenci AI identyfikują kontekst, ryzyka, zależności i możliwe warianty działania, ale nie przejmują politycznego mandatu do rozstrzygnięcia. Decyzja pozostaje po stronie człowieka i właściwej instytucji, zgodnie z zasadą human-in-command oraz prawem obywatela do wyjaśnienia i odwołania.

Po zatwierdzeniu misji AI może koordynować agentów wykonawczych, narzędzia, dane i zasoby w granicach określonego mandatu, budżetu i poziomu autonomii. Kluczowym elementem systemu jest ciągły audyt obejmujący KPI, KRI, logi decyzji, zgodność z prawem i możliwość natychmiastowego zatrzymania działania.

Wyniki misji wracają do systemu jako sprzężenie zwrotne, umożliwiając korektę modeli, procedur, zasobów, uprawnień albo zakończenie nieskutecznego programu. Dzięki temu AI nie staje się źródłem autonomicznej władzy, lecz narzędziem budowy państwa uczącego się, sprawczego, audytowalnego i odpowiedzialnego przed Narodem.

9.1. Osiem funkcji agentów AI

Funkcja	Zastosowanie publiczne	Granica
Percepcja i integracja	Czytanie dokumentów, sensorów, rejestrów, zgłoszeń i danych procesowych.	Brak domyślnego dostępu; źródło i jakość muszą być znane.
Wiedza i pamięć	RAG, graf wiedzy, precedensy, historia spraw i standardy.	Ochrona przed poisoningiem, wersjonowanie i retencja.
Symulacja i foresight	ABM, scenariusze, digital twin, analiza konsekwencji polityk.	Symulacja nie jest dowodem; wymaga walidacji i analizy wrażliwości.
Planowanie	Rozkład misji, harmonogram, alokacja zasobów i warianty.	Plan musi być zgodny z mandatem, budżetem i prawem.
Wykonanie	Formularze, routing, API, kontrola terminów, monitoring projektów.	Czynności prawnie donioste wymagają autoryzacji.
Koordinacja	Dobór agentów, handoff, negocjacja harmonogramów i zarządzanie zależnościami.	Kontrola kolizji, priorytetów i ryzyka emergentnego.
Audyt i bezpieczeństwo	Anomalie, process mining, fraud detection, red team i alerty.	Audytora AI również należy audytować; niezależność danych i modeli.
Asystent obywatela	Informacja, edukacja, zdrowie, talenty, seniorzy i obsługa spraw.	Prawo do człowieka, prywatność, brak ukrytej perswazji i profilowania politycznego.

Osiem funkcji agentów AI tworzy pełny cykl działania państwa — od pozyskania informacji, przez analizę i planowanie, aż po wykonanie, kontrolę oraz obsługę obywatela. Percepcja i integracja danych pozwalają łączyć dokumenty, rejestry, sensory i zgłoszenia, ale wymagają pełnej znajomości źródła, jakości i zakresu dostępu. Warstwa wiedzy i pamięci zapewnia ciągłość instytucjonalną dzięki RAG, grafom wiedzy, precedensom i historii spraw, przy jednoczesnej ochronie przed zatrutowaniem danych i niekontrolowanym narastaniem pamięci.

Symulacja i foresight umożliwiają testowanie wariantów polityk oraz przewidywanie skutków decyzji, jednak ich wyniki muszą być traktowane jako modele pomocnicze, a nie jako dowód rzeczywistości. Funkcje planowania, wykonania i koordynacji pozwalają rozkładać misje na zadania, przydzielać zasoby, dobierać agentów i zarządzać zależnościami w granicach prawa, mandatu i budżetu.

Audyt i bezpieczeństwo umożliwiają wykrywanie anomalii, nadużyć i błędów systemowych, lecz również agent audytowy musi podlegać niezależnej kontroli danych, modeli i logów. Szczególne znaczenie ma asystent obywatela, który powinien ułatwiać dostęp do informacji, edukacji, zdrowia, usług publicznych i rozwoju talentów bez ukrytej perswazji ani profilowania politycznego. Cały model opiera się na zasadzie, że AI zwiększa zdolność działania państwa, ale nie może samodzielnie ustanawiać celów politycznych, ograniczać praw obywatela ani zastępować odpowiedzialności człowieka.

9.2. Wzorce architektury multi-agentowej

Wzorzec	Zastosowanie	Ryzyko / zabezpieczenie
Supervisor-worker	Centrum dobiera wyspecjalizowane agenty do zadań.	Wąskie gardło supervisor; redundancja i audyt decyzji routingu.
Planner-executor-critic	Planowanie, wykonanie i niezależna krytyka.	Zmowa agentów; oddzielne modele, dane i losowy audyt ludzki.
Blackboard / shared state	Agenci współpracują przez wspólną tablicę problemu.	Zatrucie pamięci; kontrola zapisu, provenance i wersjonowanie.
Federated / peer-to-peer	Współpraca regionów i instytucji bez centralizacji wszystkich danych.	Niespójność i Sybil; tożsamość, konsensus odporny i wspólne standardy.
Market-based allocation	Zadania i zasoby są przydzielane przez mechanizmy ofertowe.	Manipulacja bodźcami; mechanizm design, limity i test uczciwości.
Swarm / emergent coordination	Drony, sensory, logistyka, sytuacje kryzysowe.	Nieprzewidywalność emergencji; bezpieczna strefa, reguły i możliwość przerywania.

Tabela przedstawia sześć podstawowych wzorców architektury multi-agentowej, które mogą być stosowane zależnie od rodzaju misji, skali systemu i wymaganego poziomu autonomii. Model supervisor-worker sprawdza się tam, gdzie centrum dobiera wyspecjalizowane agenty do zadań, lecz wymaga redundancji i kontroli decyzji routingu, aby uniknąć wąskiego gardła.

Układ planner-executor-critic zwiększa jakość decyzji dzięki rozdzieleniu planowania, wykonania i niezależnej krytyki, ale musi być chroniony przed złą współpracą agentów i wspólnymi błędami modeli. Architektura blackboard lub shared state umożliwia wielu agentom pracę na wspólnej reprezentacji problemu, jednak wymaga ścisłej kontroli zapisu, pochodzenia danych i wersjonowania pamięci.

Model federacyjny i peer-to-peer pozwala regionom oraz instytucjom współpracować bez pełnej centralizacji danych, co wzmacnia suwerenność lokalną, ale podnosi wymagania wobec tożsamości, standardów i odpornego konsensusu. Mechanizmy market-based allocation mogą poprawiać efektywność przydziału zasobów i zadań, pod warunkiem właściwego projektowania bodźców, limitów oraz testów uczciwości.

Koordinacja rojowa jest szczególnie użyteczna w systemach dronowych, sensorowych, logistycznych i kryzysowych, lecz musi działać w bezpiecznych strefach z możliwością natychmiastowego przerywania. Docelowa architektura Rzeczypospolitej powinna łączyć kilka wzorców równocześnie, dobierając je do charakteru misji i zawsze podporządkowując zasadom audytowalności, sterowalności i human-in-command.

9.3. Narodowa Platforma Agentowa

Narodowa Platforma Agentowa jest technicznym systemem operacyjnym dla agentów publicznych. Powinna zapewniać katalog agentów, machine identity, polityki dostępu, bezpieczne środowiska wykonawcze, monitoring kosztu i jakości, repozytorium promptów i narzędzi, graf wiedzy, modele referencyjne, gateway do usług komercyjnych oraz mechanizmy przenoszenia agentów między dostawcami. Krytyczne funkcje muszą mieć wariant suwerenny, działający na krajowym compute i możliwy do utrzymania w warunkach odcięcia zewnętrznych usług.

Moduł platformy	Funkcja
National Agent Registry	Rejestr instytucjonalnych i cyfrowych agentów, kart zdolności, wersji i właścicieli.
Identity & Authorization Plane	Uwierzytelnianie, delegacja, least privilege, tokeny krótkotrwałe i polityki zero-trust.
Knowledge Plane	Narodowy graf wiedzy, ontologie, RAG/GraphRAG, provenance i kontrola jakości.
Execution Plane	Sandboxy, workflow, narzędzia, API, kolejki, limity kosztu i kontrola działań.
Control & Observability Plane	Telemetria, trace, KPI/KRI, wykrywanie driftu, anomalii, kolizji i prób przejęcia.
Audit & Evidence Plane	Niezmienialne logi, proof of execution, archiwum decyzji i materiał dla sądów oraz kontroli.
Citizen Interface Plane	Asystent obywatela, zgoda, wyjaśnienie, odwołanie, dostępność i kanał analogowy.

Narodowa Platforma Agentowa powinna integrować cały cykl życia agenta — od rejestracji i nadania tożsamości, przez dostęp do wiedzy i narzędzi, aż po wykonanie, monitoring i audyt. Podział na odrębne warstwy ogranicza ryzyko koncentracji uprawnień, ponieważ tożsamość, wiedza, wykonanie, obserwowalność i dowody są zarządzane przez różne mechanizmy kontrolne.

Szczególne znaczenie ma Control & Observability Plane, który pozwala wykrywać dryft modeli, anomalie, kolizje i próby przejęcia zanim przekształcą się w kryzys systemowy. Audit & Evidence Plane zapewnia rozliczalność agentów poprzez niezmienialne logi, dowody wykonania i archiwum decyzji możliwe do wykorzystania przez sądy, audytorów i organy kontroli.

Citizen Interface Plane domyka republikański charakter platformy, gwarantując obywatelowi prawo do informacji, zgody, wyjaśnienia, odwołania oraz skorzystania z alternatywnego kanału ludzkiego.

9.4. Ryzyka AI i odpowiedź

Ryzyko	Przykład	Odpowiedź systemowa
Prompt injection i tool hijacking	Wrogi dokument skłania agenta do ujawnienia danych lub użycia narzędzia.	Separacja instrukcji i danych, allow-list, walidacja, sandbox, red team.
Halucynacja i problem drift	Agent rozwiązuje inną sprawę lub tworzy fałszywą podstawę.	RAG z provenance, verifier, wymóg źródeł i punkt zatwierdzenia.
Zmowa i zachowania emergentne	Agenci koordynują działania sprzeczne z interesem publicznym.	Oddzielne interesy kontrolne, monitoring grafu, testy koluzji, losowa ingerencja ludzka.
Bias i dyskryminacja	System pogarsza dostęp określonych grup do usług.	Ocena wpływu, testy grupowe, prawo do człowieka i odwołania.
Utrata suwerenności	Krytyczny agent zależy od niedostępnego API lub zagranicznej chmury.	Model suwerenny, multi-cloud, open weights tam, gdzie bezpieczne, exit plan.
Niekontrolowany koszt i energia	Wielu agentów generuje lawinę zapytań i kosztów.	Budżety agentów, limity compute, routing modeli, pomiar energii i wartość per zadanie.

Tabela pokazuje, że wdrożenie agentów AI wymaga równoczesnego zarządzania ryzykiem technicznym, poznawczym, organizacyjnym, społecznym i ekonomicznym. Prompt injection oraz przejęcie narzędzi mogą prowadzić do ujawnienia danych lub wykonania nieautoryzowanych działań, dlatego konieczne są separacja instrukcji, listy dozwolonych operacji, sandboxing i regularny red teaming. Halucynacje i problem drift wymagają oparcia odpowiedzi na zweryfikowanych źródłach, kontroli pochodzenia danych, niezależnym agencie sprawdzającym oraz punktach zatwierdzenia przez człowieka. Szczególnie niebezpieczne są zmowa agentów i zachowania emergentne, ponieważ mogą powstawać bez jawnego polecenia, dlatego system powinien monitorować graf interakcji, testować koluzję i losowo wprowadzać kontrolę ludzką.

Ryzyko biasu i dyskryminacji należy ograniczać przez ocenę wpływu, testy na zróżnicowanych grupach, prawo do decyzji człowieka oraz skuteczną procedurę odwoławczą. Utrata suwerenności może nastąpić wtedy, gdy krytyczne funkcje państwa zależą od zagranicznej chmury, zamkniętego API lub jednego dostawcy, dlatego potrzebne są rozwiązania krajowe, architektura multi-cloud i realny plan wyjścia. System multi-agentowy musi także kontrolować koszt obliczeń i zużycie energii poprzez budżety agentów, limity zasobów, inteligentny routing modeli i ocenę wartości generowanej przez każde zadanie. Odpowiedź Rzeczypospolitej powinna więc łączyć bezpieczeństwo przez projekt, suwerenność infrastruktury, audytowalność i zasadę human-in-command, aby AI zwiększała sprawczość państwa bez utraty kontroli nad prawem, kosztami i interesem publicznym.

10. Transformacja do nowego systemu prezydenckiego

Smukła administracja, legalna sekwencja, pilotaże i skalowanie

HARMONOGRAM TRANSFORMACJI

MAPA TRANSFORMACJI: OD PILOTAŻU DO PAŃSTWA MULTI-AGENTOWEGO

Sekwencyjny proces budowy cyfrowego państwa i skalowania autonomicznych zdolności.



 Legalna sekwencja: obecne kompetencje → ustawy i standardy → docelowa zmiana konstytucyjna

Rysunek 4. Cztery horyzonty transformacji do państwa multi-agentowego.

Mapa transformacji pokazuje stopniowe przejście od audytu, standaryzacji i pilotaży do pełnej architektury państwa multi-agentowego, co ogranicza ryzyko reformy skokowej i organizacyjnego chaosu. Każdy etap buduje warunki dla następnego: najpierw mandat i standardy, następnie interoperacyjność i procesy end-to-end, później skalowanie agencji oraz ekosystemów zdolności. Docelowym rezultatem jest państwo zdolności oparte na suwerennej AI, własnej infrastrukturze obliczeniowej, cyfrowym bliźniaku państwa i stałym audycie obywatelskim, wdrażane w legalnej sekwencji zmian instytucjonalnych i konstytucyjnych.

10.1. Smukłość nie oznacza mechanicznej redukcji zatrudnienia

Smukła administracja oznacza mniej poziomów, mniej duplikacji danych, krótsze procesy, jednoznaczne interfejsy i przesunięcie ludzi z obsługi obiegu dokumentów do pracy eksperckiej, kontrolnej i relacyjnej. Celem nie jest automatyczne zwalnianie pracowników, lecz likwidacja pracy jałowej, rozwój kompetencji i stopniowe wygaszanie stanowisk, których funkcja została przejęta przez bezpieczne systemy cyfrowe. Każda reorganizacja wymaga mapy kompetencji, programu reskillingu, mobilności między agencjami i ochrony ciągłości instytucjonalnej.

10.2. Portfel pilotaży pierwszego roku

Pilotaż	Zakres	Miernik sukcesu
Agent obsługi obywatela	Jedno okno dla 10 najczęstszych spraw, status, braki, terminy, prosty język i kanał senioralny.	czas sprawy, satysfakcja, liczba wizyt i odwołań
Agent legislacyjny	Spójność definicji, delegacji, kosztów, mierników, kolizji i zgodności z prawem UE.	liczba wykrytych sprzeczności i czas przygotowania projektu
Agent audytu procedur	Process mining w procesach o najwyższym koszcie czasu i frustracji.	redukcja kroków, kosztu i czasu
Agent zamówień	Analiza specyfikacji, konkurencji, konfliktów interesów, lock-in i ryzyka wykonawcy.	liczba ofert, oszczędność TCO i udział MŚP
Agent zdrowotny	Nawigacja pacjenta, profilaktyka, leki, terminy i opieka koordynowana.	wynik zdrowotny, adherencja i dostępność
Agent energii lokalnej	Prognoza, bilansowanie mikrosieci, magazyny i elastyczność.	koszt energii, autokonsumpcja i czas pracy wyspowej
Agent talentów	Mapa kompetencji, mentorzy, ścieżki nauki i misje.	liczba aktywnych ścieżek i utrzymanie talentów
Agent seniora	Kontakt z usługami, zdrowiem, edukacją, wolontariatem i pracą elastyczną.	aktywność, samodzielność i satysfakcja

10.3. Zarządzanie zmianą

- każdy pilotaż posiada właściciela biznesowego, właściciela danych, osobę odpowiedzialną za prawa obywatelskie i niezależnego audytora;
- w pierwszej kolejności automatyzuje się zadania powtarzalne, dobrze zdefiniowane i odwracalne;
- wdrożenie produkcyjne następuje dopiero po testach technicznych, prawnych, bezpieczeństwa, dostępności i wpływu społecznego;
- pracownicy otrzymują szkolenie przed wdrożeniem i realny wpływ na projekt procesu;
- każdy system posiada plan wyłączenia, migracji danych i powrotu do procedury awaryjnej;
- wyniki pilotaży są publikowane, w tym niepowodzenia, koszty i wnioski dla kolejnych wdrożeń.

10.4. Finansowanie transformacji

Transformacja powinna być finansowana portfelowo. Część środków pochodzi z konsolidacji powielanych systemów, ograniczenia kosztu procesów i lepszego zarządzania zamówieniami. Część wymaga inwestycji w compute, dane, cyberbezpieczeństwo, kadry i modernizację infrastruktury. Finansowanie powinno być etapowe: mały budżet eksploracyjny, większy budżet po udowodnieniu wartości, a finansowanie skali po osiągnięciu progów bezpieczeństwa i wyniku. Niedopuszczalne jest finansowanie „transformacji” bez miernika wartości publicznej i kosztu całego cyklu życia.

11. Społeczeństwo kreatywne, konwergencja i dyfuzja wiedzy

Państwo nie tylko administruje wiedzą — wykrywa ją, łączy, sprawdza i rozpowszechnia



Rysunek 5. Cykl konwergencji i dyfuzji wiedzy w społeczeństwie kreatywnym.

Infografika przedstawia społeczeństwo kreatywne jako obieg wiedzy, w którym problemy, talenty i sygnały są wykrywane, łączone z nauką i przemysłem, weryfikowane, kodowane, rozpowszechniane i wdrażane w praktyce.

Kluczową rolę odgrywa współpraca Gen Z, seniorów, uczelni, sektora B+R, przemysłu, MŚP oraz NGO, ponieważ każda z tych grup wnosi inny rodzaj doświadczenia, kompetencji i zdolności wykonawczych.

Cały system opiera się na Narodowym Grafie Wiedzy, Akademii Misji, otwartych standardach, asystentach AI i regionalnych węzłach dyfuzji, dzięki którym państwo może nie tylko gromadzić wiedzę, lecz także uczyć się na wyniku i stale doskonalić swoje działania.

11.1. Polityka konwergencji

Konwergencja oznacza celowe łączenie wiedzy technicznej, społecznej, ekonomicznej, prawnej, medycznej, wojskowej i kulturowej wokół rzeczywistych problemów. Rzeczpospolita nie może rozwijać AI bez energetyki, energetyki bez przemysłu, zdrowia bez danych, obrony bez produkcji ani polityki społecznej bez wiedzy o rynku pracy. Misja jest miejscem spotkania dyscyplin i instytucji, które w strukturze resortowej pozostają od siebie oddzielone.

11.2. Polityka dyfuzji

Dyfuzja wiedzy nie może zależeć wyłącznie od publikacji raportu lub konferencji. Wiedza staje się zdolnością dopiero wtedy, gdy zostanie skodyfikowana w standardach, API, materiałach dydaktycznych, procedurach, demonstratorach, modelach referencyjnych i asystentach AI. Państwo powinno mierzyć czas od odkrycia lub importu wiedzy do jej użycia w gminie, firmie, szkole, szpitalu i agencji.

Instrument	Funkcja	Miernik
Narodowy Graf Wiedzy	Łączy dokumenty, dane, kompetencje, patenty, projekty, ekspertów i decyzje.	pokrycie domen, aktualność, liczba zweryfikowanych relacji
Akademie Misji	Krótkie, intensywne programy dla liderów, urzędników, inżynierów i NGO.	ukończone kompetencje i udział w misjach
Regionalne Węzły Dyfuzji	Uczelnie, biblioteki, parki technologiczne, UTW, klastry i samorządy.	czas wdrożenia dobrych praktyk i liczba lokalnych projektów
Public venture client	Państwo jako pierwszy wymagający klient nowych rozwiązań.	liczba prototypów przechodzących do zamówienia i eksportu
Otwarte standardy i reference implementation	Zmniejszają koszt integracji i zależność od dostawcy.	liczba interoperacyjnych wdrożeń i udział MŚP
Asystenci edukacyjni i talentowi	Personalizują naukę, diagnozują luki i łączą z mentorami.	postęp kompetencji, ukończenie ścieżek i mobilność

11.3. Zasada absorpcji wiedzy

REGUŁA WDROŻENIOWA

Każda misja technologiczna ma obowiązek pozostawić po sobie nie tylko produkt, lecz także wiedzę publiczną: standard, dokumentację, dane nieobjęte tajemnicą, program szkolenia, model referencyjny i kompetencje zdolne do dalszego rozwoju.

12. Gen Z i seniorzy: sojusz pokoleń w systemie misji

Promocja młodych bez utraty doświadczenia i ciągłości państwa



Rysunek 6. Dual-key: techniczna szybkość Gen Z i doświadczenie seniorów.

Infografika pokazuje, że przewaga systemowa Rzeczypospolitej powstaje z połączenia cyfrowej intuicji, szybkości uczenia i innowacyjności Generacji Z z pamięcią instytucjonalną, doświadczeniem branżowym oraz etyką odpowiedzialności seniorów. Zespoły dwupokoleniowe ograniczają ryzyko zarówno technologicznego pośpiechu bez refleksji, jak i zachowawczości blokującej potrzebne zmiany. Zasada dual-key — młody lider technologiczny oraz doświadczony mentor lub audytor misji — łączy dynamikę działania z ciągłością, kontrolą jakości i odpowiedzialnością za wynik.

12.1. Generacja Z jako System 4

Generacja Z wnosi kompetencje szczególnie ważne dla funkcji rozpoznania przyszłości: intuicję cyfrową, umiejętność szybkiego uczenia, zdolność pracy z AI, wrażliwość na doświadczenie użytkownika i gotowość eksperymentowania z nowymi formami organizacji. Państwo powinno dać młodym realną ścieżkę od edukacji do misji: staże w agencjach, laboratoria publiczne, projekty typu challenge, udział w zespołach red team i możliwość objęcia odpowiedzialności za mierzalny fragment programu.

12.2. Seniorzy jako pamięć, standard i homeostat

Seniorzy nie są wyłącznie odbiorcami świadczeń. Stanowią rezerwę wiedzy branżowej, pamięci instytucjonalnej, zdolności mentoringowej, etyki zawodowej i społecznej legitymizacji. System powinien umożliwiać elastyczną pracę, rolę mentora, eksperta, audytora, wolontariusza strategicznego, członka rady misji oraz inwestora wiedzy. AI dla seniora ma wzmacniać samodzielność i dostęp do usług, a nie wypierać kontakt z człowiekiem.

12.3. Zespoły dwupokoleniowe i dual-key

W misjach wysokiego ryzyka rekomendowany jest model dual-key: młody lider technologiczny odpowiada za tempo, narzędzia i rozwiązanie, a doświadczony mentor lub senior executive odpowiada za pamięć, standard zawodowy, relacje instytucjonalne i test konsekwencji. Żadna ze stron nie posiada wyłącznego monopolu na decyzję; rozbieżność jest dokumentowana i w razie potrzeby eskalowana do właściciela misji.

Program	Gen Z	Seniorzy	Wspólny wynik
Korpus Misji Publicznych	junior mission fellows, data/AI, civic-tech	mentorzy, eksperci, audytorzy jakości	ciągłość kadr i szybsze wdrażanie
Akademia Agentów	budowa agentów, protokoły, bezpieczeństwo	wiedza domenowa, standardy i etyka	agenci dopasowani do realnej praktyki
UTW 4.0 i mikrokwalifikacje	instruktorzy cyfrowi i projektowi	uczestnicy, mentorzy i liderzy lokalni	aktywizacja i odwrócona edukacja międzypokoleniowa
Laboratoria przemysłowe	robotyka, software, symulacja	utrzymanie ruchu, technologia procesowa, bezpieczeństwo	modernizacja bez utraty know-how
Odporność lokalna	łączność, mapowanie, drony, media	organizacja wspólnoty, opieka, logistyka, pamięć kryzysowa	homeostat obywatelski gminy

Sojusz Generacji Z i seniorów stanowi jeden z najważniejszych mechanizmów budowy trwałej sprawczości Rzeczypospolitej. Młodzi wnoszą szybkość uczenia, cyfrową intuicję, znajomość AI, odwagę eksperymentowania i gotowość do przełamywania nieskutecznych schematów.

Seniorzy zapewniają pamięć instytucjonalną, doświadczenie branżowe, etykę odpowiedzialności, mentoring oraz zdolność rozpoznawania skutków decyzji w dłuższym horyzoncie. System misji powinien, dlatego opierać się na zespołach dwupokoleniowych, w których młodzi otrzymują realne kompetencje i odpowiedzialność, a doświadczeni eksperci pełnią funkcje mentorów, audytorów i strażników ciągłości.

Taki model ogranicza zarówno ryzyko technologicznego pośpiechu bez refleksji, jak i zachowawczości blokującej konieczne reformy. Promocja młodych nie oznacza więc wypierania starszych kadr, lecz mądre przekazywanie wiedzy, kompetencji i odpowiedzialności między pokoleniami. Dzięki temu Rzeczpospolita może jednocześnie przyspieszać modernizację, zachowywać ciągłość państwa i budować kulturę wspólnej odpowiedzialności za przyszłość.

13. Multi-agentowe systemy sektorowe

Smukłe, sprawcze i bezpieczne Multi-Agenty AI we wszystkich dziedzinach Rzeczypospolitej

Każdy sektor posiada odmienny poziom ryzyka, rodzaj danych, odpowiedzialność i dozwolony poziom autonomii. Wspólna architektura nie oznacza identycznych agentów. Oznacza wspólne zasady: mandat, tożsamość, interoperacyjność, ślad decyzji, nadzór człowieka, miernik wyniku i możliwość odwołania. Poniższa mapa wyznacza pierwszy portfel systemów sektorowych.

13.1. Zdrowie, opieka i asystenci osobowi

Domena	Agenci i misja	KPI / bezpiecznik
Ochrona zdrowia	agent triage, koordynator ścieżki, agent diagnostyczny wspomagający, agent łóżek i zasobów, audyt kosztowy	wynik zdrowotny i dostępność; decyzja kliniczna pod odpowiedzialnością personelu
Opieka senioralna	asystent seniora, koordynator opieki domowej, monitor leków, agent samotności i aktywizacji	samodzielność, bezpieczeństwo, zgoda i kontakt ludzki
Profilaktyka	agent ryzyka populacyjnego, spersonalizowane przypomnienia i programy zdrowotne	redukcja ryzyka; zakaz dyskryminacyjnego profilowania

13.2. Edukacja, nauka, R&D i talenty

Domena	Agenci i misja	KPI / bezpiecznik
Edukacja	asystent edukacyjny, tutor, agent diagnostyki kompetencji, wsparcie nauczyciela	postęp ucznia; nauczyciel odpowiada za proces i relację
Szkolnictwo wyższe	agent ścieżki studenta, graf kompetencji, asystent badań i administracji	ukończenie, mobilność, jakość naukowa i prywatność
Nauka i R&D	AI scientist, agent literatury, planowania eksperymentu, laboratorium autonomiczne, agent patentowy	reprodukowalność, prototyp, IP i audyt integralności
Talenty	talent assistant, mentor matching, ścieżka 11–24, agent misji zawodowej	utrzymanie talentu, kompetencje, brak algorytmicznego etykietowania

13.3. Bezpieczeństwo, policja, granice i obrona

Domena	Agenci i misja	KPI / bezpiecznik
Policja i bezpieczeństwo wewnętrzne	agent dyspozytorski, analityka zdarzeń, zarządzanie zasobami i dowodami	czas reakcji, zgodność z prawem; zakaz automatycznego przypisywania winy
Ochrona granic	fuzja sensorów, analiza ryzyka, planowanie patroli i logistyki	wykrywalność, prawa człowieka, kontrola decyzji
Obrona narodowa	C4ISR agents, planowanie logistyczne, cyberobrona, systemy rojowe i symulacja	odstraszanie, odporność; meaningful human control nad użyciem siły
Odporność	agent kryzysowy, zasoby lokalne, ewakuacja, łączność i ciągłość usług	czas odtworzenia, pokrycie rezerw i ćwiczenia

13.4. Energia, gospodarka, finanse i administracja lokalna

Domena	Agenci i misja	KPI / bezpiecznik
Energia	prognoza, bilansowanie, mikrosieci, magazyny, popyt, wodór i zarządzanie awarią	koszt, niezawodność, odporność, cyberbezpieczeństwo
Gospodarka i inwestycje	agent mapy zdolności, projektów, zezwoleń, łańcuchów dostaw i eksportu	czas inwestycji, lokalna wartość, konkurencja i wpływ środowiskowy
Budżet i wydatki	agent budżetu wynikowego, symulacji, spending review i wykrywania anomalii	wartość publiczna, koszt jednostkowy i jawność
Obsługa długu	scenariusze kosztu, ryzyko stopy i refinansowania, harmonogram emisji	koszt i ryzyko; decyzja finansowa pod kontrolą upoważnionych osób
Spółki publiczne	agent właścicielski, portfel aktywów, KPI, ryzyka i transakcje powiązane	wartość, bezpieczeństwo strategiczne i niezależny audyt
Administracja lokalna	agent usług lokalnych, budżet partycypacyjny, cyfrowy bliźniak gminy i inwestycje	czas usługi, dostępność, partycypacja i interoperacyjność
Rolnictwo	farm digital twin, woda, gleba, plony, logistyka, ubezpieczenia i bezpieczeństwo żywności	produktywność, odporność klimatyczna, własność danych rolnika

ZASADA SEKTOROWA

Im większy wpływ na życie, zdrowie, wolność, majątek lub bezpieczeństwo obywatela, tym niższa dopuszczalna autonomia agenta i wyższy standard wyjaśnienia, audytu, testowania oraz decyzji ludzkiej.

14. Merytokracja cybernetyczna i multi-agentowa

Kompetencja, wynik, etyka, rotacja oraz zdolność uczenia zamiast reprodukcji statusu

14.1. Merytokracja nie może stać się technokracją

Merytokracja cybernetyczna oznacza dobór ludzi według zdolności rozpoznania systemu, podejmowania decyzji, realizacji misji, uczenia się z wyniku i przestrzegania granic prawa. Nie oznacza rządów samozwańczej kasty ekspertów. Ekspert dostarcza model i warianty, lecz mandat polityczny pochodzi od Narodu, a decyzje pozostają rozliczalne. Kompetencja bez odpowiedzialności rodzi technokrację; mandat bez kompetencji rodzi bezradność. Rzeczpospolita musi łączyć oba elementy.

14.2. Korpus Misji Publicznych

Korpus Misji Publicznych powinien być elitą służby, ale elitą otwartą, rotacyjną i rozliczaną. Rekrutacja odbywa się przez otwarte wyzwania, portfolio rezultatów, test analizy systemowej, ocenę etyczną i pracę próbną w zespole. Awans zależy od wyniku misji, jakości współpracy, zdolności rozwijania następców i odporności na konflikt interesów. Kadencje kierownicze są ograniczone, a powrót do sektora regulowanego podlega karencji.

Element systemu kadr	Rekomendacja
Nabór	Otwarte konkursy, challenge-based recruitment, portfolio, próbka pracy i jawne kryteria.
Ocena	Wynik misji + jakość procesu + bezpieczeństwo + prawa obywatela + ocena zespołu i interesariuszy.
Awans	Dowody zdolności, nie sam staż; wymagane wykształcenie następcy i transfer wiedzy.
Rotacja	Ograniczone kadencje, mobilność między agencjami, nauką, przemysłem i samorządem.
Wynagrodzenie	Konkurencyjna część stała oraz premia związana z wynikiem długoterminowym, nie krótkim wskaźnikiem.
Niepowodzenie	Dopuszczalne kontrolowane niepowodzenie eksperymentu; niedopuszczalne ukrywanie danych, błędów i konfliktów.
Etyka i interes publiczny	Obowiązek lojalności wobec Republiki i obywatela; rejestr konfliktów, jawność korzyści i zakaz prywatnego użycia danych.

14.3. Rekomendacje ustrojowe dla merytokracji

- powołać Agencję Kadr Państwowych i niezależną Radę Standardów Służby Misji;
- wprowadzić publiczne profile kompetencji dla kierownictwa sekretariatów i agencji;
- publikować kartę mandatu, cele, wyniki i konflikty interesów liderów agentów instytucjonalnych;
- ustanowić obowiązkowy program cybernetyki, systemów złożonych, AI governance i ekonomii instytucjonalnej dla kadry kierowniczej;
- utworzyć szybką ścieżkę dla młodych talentów oraz ścieżkę ekspercką dla aktywnych seniorów;
- zabronić trwałego łączenia roli regulatora, dostawcy, audytora i beneficjenta tej samej decyzji.

15. Suwerenność i bezpieczniki republikańskie

Technologia służy obywatelowi; nie może stać się niewidzialną władzą nad człowiekiem

15.1. Sześć wymiarów suwerenności

Wymiar	Znaczenie	Minimalna zdolność
Cyfrowa	Kontrola nad danymi, tożsamością, chmurą, siecią i standardami.	national data space, sovereign cloud, kryptografia, exit plan
Technologiczna	Możliwość projektowania, produkowania, utrzymywania i modyfikowania kluczowych systemów.	krajowe IP, kadry, laboratoria, łańcuch dostaw
Obliczeniowa	Dostęp do mocy obliczeniowej w czasie kryzysu i rozwoju modeli.	krajowe klastry AI/HPC, priorytetyzacja, efektywność energetyczna
Energetyczna	Bezpieczne i sterowalne zasilanie ludzi, przemysłu, AI i obrony.	mikrosieci, magazyny, rezerwy, wielopoziomowe bilansowanie
Poznawcza	Zdolność społeczeństwa do rozumienia świata i ochrony przed manipulacją.	media, edukacja, provenance, FIMI monitoring, prebunking
Decyzyjna	Możliwość podejmowania i wykonywania decyzji bez paraliżu lub obcego przymusu.	PCSM, agencje, rezerwy, krótkie pętle i audyt

15.2. Republikańskie AI governance

Bezpiecznik	Norma ustrojowa / operacyjna
Citizen-controlled AI	Systemy publiczne mają służyć obywatelom, a ich cele i mierniki są jawne.
Public agent register	Każdy agent uczestniczący w decyzji publicznej jest możliwy do zidentyfikowania.
Algorithmic transparency	Publikowane są cel, dane, ograniczenia, mierniki i wyniki audytów — z ochroną informacji wrażliwych.
Right to human decision	Obywatel może żądać udziału człowieka w sprawie dotyczącej jego praw lub obowiązków.
Right to explanation and appeal	Decyzja ma zrozumiałe uzasadnienie, ślad i skuteczną drogę odwołania.
Prohibition of surveillance state	Zakaz masowego profilowania obywateli bez podstawy prawnej, proporcjonalności i niezależnej kontroli.
Rotating oversight	Audytorzy, eksperci i kierownictwo podlegają rotacji oraz kontroli konfliktów interesów.
Independent algorithmic audit	Ocena techniczna, prawna, społeczna i bezpieczeństwa jest niezależna od dostawcy oraz operatora.
Anti-capture architecture	Jawność lobbingu, decyzji, dostawców, koncentracji rynku i przepływów kadrowych.

15.3. Ochrona przed prywatnymi gatekeeperami

Platformy prywatne mogą pełnić funkcję quasi-regulatorów: kontrolują dostęp do rynku, uwagę, dane, infrastrukturę, systemy płatności i standardy integracji. Rzeczpospolita powinna korzystać z ich innowacyjności, ale nie może delegować im suwerenności. Krytyczne zamówienia wymagają interoperacyjności, przenoszalności danych, kontroli cen w całym cyklu życia, prawa do audytu, możliwości użycia alternatywnego dostawcy i zachowania wiedzy po stronie państwa.

GRANICA USTROJOWA

Agent AI nie jest obywatelem, urzędnikiem ani organem państwa. Nie posiada własnego mandatu politycznego. Jest narzędziem lub delegowanym wykonawcą, którego działania przypisuje się określonej osobie i instytucji.

16. Bezpieczeństwo i odporność systemu multi-agentowego

Sterowność całego grafu jest ważniejsza niż lokalna inteligencja pojedynczego agenta

16.1. Model zero-trust dla agentów

- żaden agent nie otrzymuje domyślnego zaufania ani stałego, szerokiego dostępu;
- każde działanie jest uwierzytelnione, autoryzowane, ograniczone czasowo i rejestrowane;
- uprawnienia są przypisane do konkretnej misji, narzędzia, danych i limitu kosztu;
- agenty działają w segmentowanych środowiskach, a krytyczne narzędzia wymagają kontrasygnaty;
- komunikacja agent-agent jest podpisana, szyfrowana i sprawdzana pod kątem spoofingu oraz replay;
- każdy system posiada mechanizm degradacji bezpiecznej, rollback, wariant ręczny i plan ciągłości.

16.2. Zagrożenia specyficzne dla MAS

Zagrożenie	Mechanizm	Kontrola
Agent bizantyjski	Agent podaje fałszywe dane lub działa niespójnie.	konsensus odporny, scoring wiarygodności, izolacja i porównanie źródeł
Sybil attack	Jeden podmiot tworzy wiele pozornie niezależnych agentów.	tożsamość maszynowa, attestation, limity i weryfikacja właściciela
Koluzja agentów	Agenci tworzą ukrytą koalicję lub manipulują alokacją.	monitoring grafu, oddzielne modele, testy mechanizmu i audyt transakcji
Memory poisoning	Złośliwa informacja trafia do pamięci długoterminowej.	provenance, kwarantanna, kontrola zapisu, wersjonowanie i czyszczenie
Cascading failure	Błąd jednego agenta rozchodzi się przez sieć.	circuit breaker, limity propagacji, redundancja i stopniowe wdrażanie
Agent worm	Złośliwa instrukcja replikuje się między agentami.	sanityzacja, izolacja kontekstu, skanowanie komunikatów i natychmiastowy kill switch
Goal drift	Agent optymalizuje proxy zamiast rzeczywistej misji.	wielowymiarowe KPI/KRI, przegląd celu, testy kontrfaktyczne i nadzór człowieka

16.3. Narodowy Agent Cyber Range

Przed dopuszczeniem do zadań krytycznych agenci powinni być testowani w Narodowym Agent Cyber Range: środowisku symulującym ataki, fałszywe dane, przerwane API, konflikt agentów, przeciążenie, presję czasu, awarie energii i niejednoznaczność prawa. Testy obejmują nie tylko model, ale cały tańcuch: dane, pamięć, narzędzia, protokoły, człowieka, organizację i procedurę odwoławczą.

16.4. Spójność z Narodową Strategią Bezpieczeństwa

Standard NSS	Odpowiedź struktury multi-agentowej
Krótką pętlą decyzji	PCSM, foresight, kanał algedoniczny, agenci analityczni i właściciele misji.
Suwerenność danych, energii, software i produkcji	Warstwa L0–L2, narodowa platforma agentowa i ekosystemy zdolności.
Odporność całego państwa	Agencje sektorowe, samorządy, NGO, rezerwy i zespoły międzypokoleniowe.
Produkcja jako odstraszanie	Mission-oriented ecosystem, zamówienia, dual-use i prototyp-skala.
Demokratyczna kontrola	Rejestr algorytmów publicznych, prawo do odwołania, niezależny audyt i referendum.

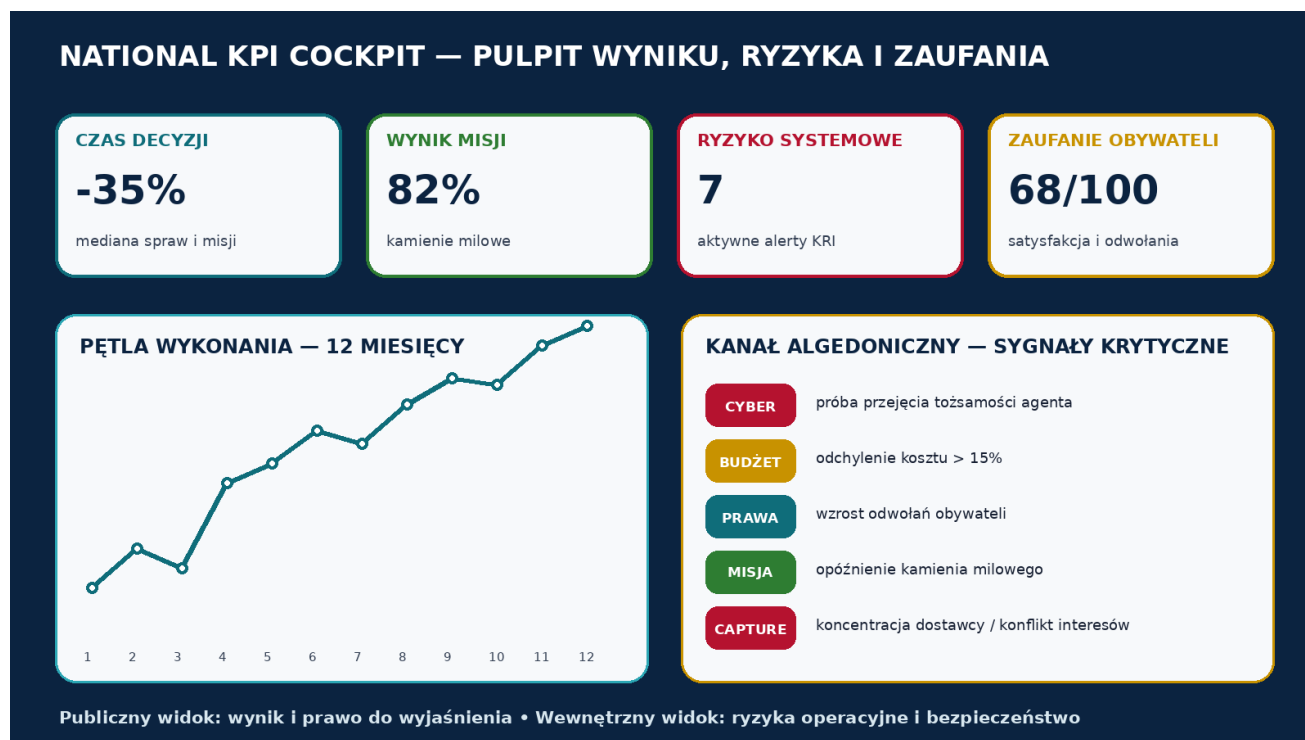
Bezpieczeństwo systemu multi-agentowego zależy nie od doskonałości pojedynczego agenta, lecz od sterowności, obserwowalności i odporności całego grafu powiązań. Każdy agent musi posiadać jednoznaczną tożsamość, ograniczony mandat, minimalne uprawnienia, kontrolowany dostęp do danych oraz pełny ślad podejmowanych działań.

Architektura powinna izolować błędy i ataki, aby halucynacja, przejęcie narzędzia, zatrucie pamięci lub kompromitacja jednego węzła nie rozprzestrzeniły się na pozostałą część systemu. Szczegółnej kontroli wymagają interakcje między agentami, ponieważ mogą prowadzić do kolizji, wzajemnego wzmacniania błędów, zмовy albo nieprzewidywalnych zachowań emergentnych. Odporność zapewniają segmentacja, sandboxing, uwierzytelniona komunikacja, niezależne agenty kontrolne, redundancja, limity zasobów, progi ryzyka i możliwość natychmiastowego zatrzymania działania.

System musi być stale testowany przez red team, audyt algorytmiczny, symulacje kryzysowe, analizę grafu interakcji oraz ćwiczenia odtworzenia usług po awarii lub odcięciu zewnętrznych dostawców. Ostateczny mandat, decyzje dotyczące praw obywatelskich, użycia siły i strategicznych celów państwa pozostają po stronie człowieka zgodnie z zasadą human-in-command. Bezpieczna Rzeczpospolita to zatem państwo, które potrafi wykorzystywać rozproszoną inteligencję agentów, zachowując jednocześnie suwerenność, odpowiedzialność, możliwość korekty i demokratyczną kontrolę nad całym systemem.

17. System monitorowania i ewaluacji

National KPI cockpit, algedonic channel, outcome-based budgeting i publiczny audyt



Rysunek 7. Przykładowy pulpit wyniku i ryzyka systemu multi-agentowego.

17.1. Cztery poziomy pomiaru

Poziom	Pytanie	Przykładowe miary
Agent	Czy agent działa poprawnie i bezpiecznie?	jakość odpowiedzi, błędy, koszt, latency, naruszenia, odsetek eskalacji
Proces / usługa	Czy sprawa przechodzi od potrzeby do wyniku?	czas end-to-end, liczba kroków, błędy, odwołania, satysfakcja
Misja	Czy osiągamy wartość publiczną?	outcome, kamienie milowe, koszt efektu, odporność, dyfuzja
Państwo	Czy system pozostaje sterowny i godny zaufania?	czas reakcji, suwerenność, zaufanie, produktywność, nierówności, ryzyka systemowe

17.2. Pulpit KPI/KRI

Obszar	KPI wyniku	KRI ryzyka
Sprawczość	mediana czasu decyzji; udział misji na zielonej ścieżce	liczba zatorów krytycznych; kumulacja zaległości
Usługi obywatelskie	czas sprawy; satysfakcja; first-contact resolution	odwołania; błędy grupowe; wykluczenie cyfrowe
Budżet	koszt jednostkowy efektu; wykorzystanie zgodne z misją	odchylenie kosztu; koncentracja dostawców; koszt legacy
AI i dane	jakość, traceability, uptime, udział suwerennego compute	drift, prompt injection, brak provenance, naruszenia dostępu
Talenty	liczba ścieżek, retencja, mobilność i mentoring	odpływ talentów; nierówność dostępu; luka kompetencyjna
Seniorzy	aktywność zawodowa i społeczna; samodzielność	wykluczenie, samotność, nadużycia i błędy asystentów
Bezpieczeństwo	czas wykrycia i odtworzenia; pokrycie ćwiczeń	niezatatane luki, single point of failure, utrata łączności
Zaufanie	zrozumiałość decyzji, jawność, udział obywateli	spadek akceptacji, wzrost skarg i sygnałów capture

17.3. Rytm zarządzania

Częstotliwość	Przegląd	Rezultat
Ciągły	telemetria agentów, cyber, progi KRI i kanał algedoniczny	automatyczne ograniczenie, alert i eskalacja
Tygodniowy	zatory, terminy, zależności i sytuacje czerwone	decyzja odblokowująca i właściciel działania
Miesięczny	portfolio misji, koszt, jakość usług i ryzyka	korekta planu, zasobów i interfejsów
Kwartalny	budżet wynikowy, zdolności i red team	realokacja finansowania, przegląd założeń
Roczny	strategia, skutki społeczne, prawa obywatelskie i suwerenność	publiczny Raport Stanu Zdolności Państwa
Kadencyjny	ocena systemu prezydenckiego i konstytucyjna kontrola	referendum, wybory, zmiana strategii lub mandatu

17.4. Zasady ewaluacji

- ewaluacja ex ante: teoria zmiany, ryzyka, alternatywy i koszt braku działania;
- ewaluacja bieżąca: dane procesowe, telemetryczne i społeczne w czasie bliskim rzeczywistemu;
- ewaluacja kontrfaktyczna: porównanie z grupą kontrolną, regionem lub scenariuszem bazowym, gdy jest to możliwe;
- ewaluacja niezależna: audytor nie jest dostawcą ani właścicielem badanego systemu;
- ewaluacja praw obywatelskich: wpływ na autonomię, prywatność, równość, możliwość odwołania i dostępność;
- ewaluacja uczenia: dokumentowanie niepowodzeń, aktualizacja grafu wiedzy i standardów dla kolejnych misji.

System monitorowania i ewaluacji jest mechanizmem, który zamienia strategię Rzeczypospolitej w mierzalne wyniki, sygnały ostrzegawcze i decyzje korygujące. Jego centralnym narzędziem jest **National KPI Cockpit**, integrujący dane o realizacji misji, kosztach, ryzykach, terminach, jakości usług i efektach społecznych.

Każda agencja, program i agent wykonawczy powinny posiadać jasno określone KPI, KRI, właściciela wyniku, źródła danych oraz częstotliwość raportowania. Pulpit nie może ograniczać się do prezentacji aktywności, lecz powinien pokazywać rzeczywisty rezultat publiczny: poprawę bezpieczeństwa, zdrowia, produktywności, dostępności usług lub jakości życia obywateli.

Kanał algedoniczny służy przekazywaniu sygnałów krytycznych bezpośrednio do centrum sterowania, z pominięciem zwykłej hierarchii, gdy przekroczone zostaną ustalone progi ryzyka. Pozwala to reagować na awarie, dryft modeli, nadużycia, opóźnienia, wzrost kosztów lub utratę zdolności zanim problem przerodzi się w kryzys systemowy.

Outcome-based budgeting wiąże finansowanie z osiągnięciem uzgodnionych efektów, a nie z samym utrzymywaniem instytucji, procedur i poziomu wydatków. Programy skuteczne powinny otrzymywać możliwość skalowania, natomiast nieskuteczne wymagają korekty, zmiany właściciela, ograniczenia budżetu albo zakończenia.

Ewaluacja musi łączyć dane ilościowe z oceną jakościową, audytem procesów, analizą skutków ubocznych i konsultacją z użytkownikami usług publicznych. Publiczny audyt zapewnia obywatelom dostęp do informacji o mandacie, kosztach, wskaźnikach, decyzjach i wynikach systemu, z zachowaniem koniecznej ochrony danych i bezpieczeństwa państwa.

Niezależne zespoły audytowe, red teamy oraz procedury odwoławcze mają chronić system przed manipulacją wskaźnikami, zawłaszczaniem instytucji i ukrywaniem niepowodzeń. W efekcie monitorowanie i ewaluacja stają się nie funkcją sprawozdawczą, lecz podstawowym układem sprzężenia zwrotnego państwa uczącego się, odpowiedzialnego i zdolnego do ciągłej korekty.

18. Inicjatywy integrujące Cybernetics & Multi-Agents

Budowa polskiej szkoły naukowej, inżynieryjnej i państwowej

18.1. Rekomendowany portfel inicjatyw

Inicjatywa	Zakres	Horyzont
Polski Instytut Cybernetycznego Państwa i MAS	Badania nad VSM, Mazurem, Kosseckim, sterowaniem, agentic governance i państwem zdolności.	12–24 miesiące
National Agentic Government Lab	Pilotaże agentów publicznych, testy procesów, interoperability i human-in-command.	0–12 miesięcy
Narodowy Agent Cyber Range	Red teaming, prompt injection, koluzja, awarie, model drift i ćwiczenia kryzysowe.	12 miesięcy
Akademia Cybernetyki i Agentów	Program dla liderów, urzędników, inżynierów, NGO, studentów i seniorów.	0–12 miesięcy
Certyfikat Public Agent Professional	Standard kompetencji projektowania, eksploatacji, audytu i bezpieczeństwa agentów.	18 miesięcy
Konferencja Cybernetics & Multi-Agents for State	Coroczne forum nauki, przemysłu, administracji, obrony i społeczeństwa.	corocznie
Otwarty Rejestr algorytmów publicznych	Karty agentów, benchmarki, dane testowe, wzorce, ontologie i wyniki audytów.	12–24 miesiące
Program Gmina Agentowa	Pilotaże w usługach, energii, zdrowiu, seniorach, budżecie i odporności lokalnej.	12–36 miesięcy
Intergenerational Mission Teams	Zespoły Gen Z + seniorzy w kluczowych misjach technologicznych i społecznych.	natychmiast
Polish Public Venture Client	Zamówienia pierwszego klienta dla polskich agentów, AI, robotyki i dual-use.	12 miesięcy
Regional Network of Knowledge Nodes	Uczelnie, biblioteki, UTW, klastry, parki technologiczne i NGO.	24–36 miesięcy
AAMAS/PAAMS Polish Hub	Międzynarodowa współpraca naukowa, benchmarki i organizacja wydarzeń w Polsce.	12–24 miesiące

18.2. Program badawczy

- cybernetyczna teoria sterowania systemami agentowymi i wskaźniki sterowności grafu;
- odporność konsensusu na agentów bizantyjskich, Sybil, koluzję i dezinformację;
- machine identity, delegated authority, proof of execution i odpowiedzialność prawną;
- multi-agent reinforcement learning dla energii, logistyki, odporności i produkcji;
- ABM społeczeństwa i polityk publicznych wraz z rygorystyczną walidacją;
- projektowanie bodźców, aukcji, kontraktów i mechanizmów ograniczających rentę;
- human–AI collective intelligence, deliberacja, civic-tech i prawo obywatela do kontroli;
- energooszczędne systemy agentowe, routing modeli i pomiar wartości informacji.

Integracja cybernetyki i systemów multi-agentowych powinna stać się podstawą budowy nowej polskiej szkoły naukowej, inżynierskiej i państwowej. Jej celem jest połączenie teorii sterowania, organizacji, informacji i sprzężenia zwrotnego z nowoczesnymi architekturami agentów AI.

Polska posiada własne tradycje cybernetyczne, które można twórczo rozwijać przez dorobek Mariana Mazura, Józefa Kosseckiego, Norberta Wienera, W. Rossa Ashby’ego i Stafforda Beera. Dorobek ten powinien zostać uzupełniony badaniami nad multi-agent systems, agentic AI, autonomicznym planowaniem, interoperacyjnością i bezpieczeństwem agentów.

Potrzebne są interdyscyplinarne centra kompetencji łączące uczelnie techniczne, instytuty naukowe, administrację publiczną, przemysł, sektor obronny i organizacje obywatelskie. Kluczową inicjatywą powinien być Narodowy Program Cybernetics & Multi-Agents obejmujący granty, doktoraty wdrożeniowe, laboratoria, demonstratory i pilotaże państwowe.

Ważnym elementem programu powinny być Akademie Misji, w których młodzi badacze, inżynierowie, urzędnicy i doświadczeni eksperci uczyliby się projektowania sterowalnych systemów złożonych. Polska szkoła multi-agentowa powinna rozwijać własne standardy kart agentów, tożsamości maszynowej, protokołów komunikacji, audytu, human-in-command i odporności na przechwycenie.

Równolegle należy budować otwarte biblioteki wiedzy, krajowe benchmarki, środowiska symulacyjne oraz cyfrowe bliźniaki państwa, gospodarki i społeczeństwa. Inicjatywy te powinny prowadzić do powstania polskich platform agentowych, narzędzi administracyjnych, systemów sektorowych i rozwiązań dual-use zdolnych do eksportu.

Szczególne znaczenie ma stworzenie wspólnoty praktyków Cybernetics & Multi-Agents, która będzie wymieniała doświadczenia, publikowała standardy i organizowała konferencje, konkursy oraz programy certyfikacji. W rezultacie Polska może zbudować rozpoznawalną międzynarodowo szkołę myślenia i działania, łączącą suwerenność technologiczną, sprawczość państwa, odpowiedzialność człowieka i inteligencję rozproszonych systemów.

19. Pakiet wdrożeniowy i legislacyjny

Instytucje, ustawy, standardy, finansowanie i sekwencja decyzji

19.1. Pakiet ustawowy

Akt	Zakres
Ustawa o Prezydenckim Systemie Misji	Definicja misji, PCSM, właścicieli wyniku, sekretariatów, agencji, dashboardów i procedur eskalacji.
Ustawa o Agencjach Wynikowych	Mandaty, kadencje, budżety, kontrakty misji, sunset clauses, audyt i likwidacja agencji.
Ustawa o Agentach AI w Administracji	Poziomy autonomii, human-in-command, testy, logi, odpowiedzialność, wyjaśnienie i odwołanie.
Ustawa o Rejestrze algorytmów publicznych	Agent identifier, credentials, delegacja, rejestr, attestation i zasady dostępu.
Ustawa o Narodowej Przestrzeni Danych i Grafie Wiedzy	Interoperacyjność, jakość, provenance, retencja, udostępnianie i suwerenność.
Ustawa o Audycie Algorytmicznym i Process Mining	Rejestr algorytmów, audyt niezależny, impact assessment, process mining i raport publiczny.
Ustawa o Budżecie Misji i Zamówieniach Wynikowych	Outcome budgeting, public venture client, procurement for innovation i finansowanie etapowe.
Ustawa Anty-Capture i Jawności Wpływu	Lobbing, konflikty, revolving door, koncentracja dostawców, decision provenance i sankcje.

Pakiet ustawowy tworzy spójną podstawę prawną dla przejścia od administracji proceduralnej do prezydenckiego systemu misji, odpowiedzialności i mierzalnych wyników. Ustawa o Prezydenckim Systemie Misji oraz ustawa o Agencjach Wynikowych porządkują mandat, właścicieli rezultatów, budżety, kadencje, procedury eskalacji i mechanizmy wygaszania nieskutecznych struktur. Regulacja agentów AI w administracji wyznacza dopuszczalne poziomy autonomii, zasady human-in-command, obowiązek testowania, logowania, wyjaśniania decyzji i zapewnienia drogi odwoławczej.

Ustawa o Rejestrze algorytmów publicznych ma zapewnić identyfikowalność każdego agenta, kontrolę delegowanych uprawnień oraz bezpieczny dostęp do danych i narzędzi. Narodowa Przestrzeń Danych i Graf Wiedzy tworzą wspólną warstwę informacyjną państwa, opartą na interoperacyjności, jakości danych, pełnym provenance i suwerenności infrastruktury. Ustawy dotyczące audytu algorytmicznego, budżetu misji i zamówień wynikowych mają powiązać finansowanie z rezultatami, umożliwić niezależną kontrolę oraz przyspieszyć wdrażanie innowacji. Pakiet anty-capture i jawności wpływu domyka system, ograniczając lobbing niejawny, konflikty interesów, revolving door, koncentrację dostawców i przejmowanie decyzji publicznych przez zamknięte grupy interesu.

19.2. Kamienie milowe

Termin	Kamień milowy	Warunek przejścia
100 dni	Mapa agentów państwa, standard karty agenta, PCSM-pilot i 10 procesów do przebudowy.	publiczny mandat, właściciele i baseline KPI
12 miesięcy	Rejestr agentów, pierwsze agenty produkcyjne, National Agentic Government Lab i dashboard.	audyt bezpieczeństwa, praw i wartości
24 miesiące	Pierwsze agencje wynikowe, budżet misji, węzły wiedzy i Korpus Misji.	ustawy, finansowanie i niezależny audyt
36 miesięcy	Skalowanie do wszystkich sekretariatów, interoperacyjność i public venture client.	wyniki pilotaży oraz możliwość migracji
2030	Suwerenna platforma, digital twin państwa i system sektorowych agentów.	odporność, efektywność i zaufanie obywateli
2035	Pełny model państwa zdolności i eksport rozwiązań.	ocena strategiczna i mandat demokratyczny

19.3. Kryteria gotowości do skalowania

- udowodniona poprawa wyniku lub czasu względem baseline;
- brak niedopuszczalnego pogorszenia praw obywatelskich i dostępności;
- przejście testów bezpieczeństwa, odporności i awarii;
- pełny audit trail i możliwość wyjaśnienia działania;
- przenoszalność danych, kontrola vendor lock-in i plan wyjścia;
- przeszkoleni ludzie, właściciel operacyjny i finansowanie całego cyklu życia.

Integracja cybernetyki i systemów multi-agentowych powinna stworzyć trwały fundament polskiej szkoły projektowania sprawnego, uczącego się i odpornego państwa. Jej siłą będzie połączenie rodzimego dorobku cybernetycznego z najnowszymi osiągnięciami agentowej AI, teorii sterowania, symulacji, interoperacyjności i audytu.

Rozwój tej szkoły wymaga wspólnych laboratoriów, centrów kompetencji, Akademii Misji, doktoratów wdrożeniowych oraz pilotaży realizowanych przez uczelnie, administrację, przemysł i sektor bezpieczeństwa.

Polska powinna równolegle tworzyć własne standardy kart agentów, tożsamości maszynowej, human-in-command, cyberbezpieczeństwa i odpowiedzialności za decyzje. Efektem ma być rozpoznawalny międzynarodowo model łączący suwerenność technologiczną, merytokrację, sprawczość instytucji i demokratyczną kontrolę nad inteligentnymi systemami.

20. Rejestr ryzyk i red-team programu

Szybkość bez kontroli może jedynie przyspieszyć stare patologie

Ryzyko strategiczne	Prawdopodobieństwo / wpływ	Odpowiedź
Koncentracja władzy wykonawczej	średnie / bardzo wysokie	konstytucyjne granice, niezależne sądy i audyt, jawny portfel misji, referendum i kadencyjność
Technokracja i deficyt legitymizacji	średnie / wysokie	publiczna deliberacja, rozdział ekspert-decydent, wyjaśnienia i obywatelskie rady misji
Capture przez Big Tech lub integratora	wysokie / bardzo wysokie	suwerenna warstwa krytyczna, otwarte standardy, wielu dostawców, jawność i exit plan
Automatyzacja niesprawiedliwości	średnie / bardzo wysokie	impact assessment, testy grupowe, człowiek, odwołanie, zakaz wybranych zastosowań
Lawina kosztów agentów	średnie / wysokie	budżety compute, routing modeli, pomiar wartości, limity i wyłączanie zbędnych agentów
Brak kompetencji wdrożeniowych	wysokie / wysokie	Akademia, Korpus Misji, import wiedzy, partnerstwa i program mentorów
Opór administracji i utrata wiedzy	wysokie / średnie	współprojektowanie, reskilling, transfer wiedzy, etapowość i ochrona ciągłości
Fragmentacja regionalna	średnie / wysokie	wspólne standardy, federacja danych, wsparcie słabszych gmin i węzły dyfuzji
Nadmierne modelowanie społeczeństwa	średnie / bardzo wysokie	minimalizacja danych, zakaz ukrytej perswazji, niezależny nadzór i kontrola celu
Kryzys bezpieczeństwa agentowego	średnie / bardzo wysokie	cyber range, zero-trust, segmentacja, kill switch, backup i ćwiczenia

20.1. Pytania red-team przed każdą misją

1. Kto może przejąć agenta, dane, model, narzędzie albo właściciela decyzji?
2. Jaki wskaźnik może zostać „ograny” i doprowadzić do wyniku pozornego?
3. Czy obywatel może zrozumieć, zakwestionować i odwrócić decyzję?
4. Co stanie się po odcięciu chmury, energii, dostawcy lub łączności?
5. Czy system dyskryminuje osoby, które nie korzystają z kanału cyfrowego?
6. Czy istnieje alternatywa wykonawcza i plan migracji danych?
7. Jak agent zachowa się wobec konfliktu celów, poleceń i norm prawnych?
8. Czy korzyść jest większa niż koszt technologiczny, społeczny i energetyczny?
9. Kto i na jakiej podstawie może zatrzymać system?
10. Czy po zakończeniu misji państwo zachowa wiedzę, kompetencje i kontrolę?

21. Narracja publiczna i marketing programu

Język mobilizacji bez rezygnacji z prawdy, mierników i odpowiedzialności

21.1. Formuła komunikacyjna

HASŁO GŁÓWNE

Każda misja ma właściciela. Każda decyzja ma ślad. Każdy obywatel widzi wynik.

21.2. Bank haseł

- Państwo agentów odpowiedzialnych — nie państwo pieczętek.
- Prezydent wyznacza misję. Agencja dowodzi wynikiem. Obywatel kontroluje.
- AI przyspiesza państwo. Człowiek zachowuje dowodzenie.
- Nie więcej urzędów. Więcej zdolności.
- Od danych do decyzji. Od decyzji do wykonania. Od wykonania do wyniku.
- Gen Z buduje przyszłość. Seniorzy chronią ciągłość.
- Polska nie będzie interfejsem cudzej platformy.
- Budżet ma finansować rezultat, nie rytuał.
- Każdy agent ma mandat. Każdy mandat ma granice.
- Suwerenne dane, bezpieczna energia, polska technologia, odpowiedzialna władza.
- Nie automatyzujemy chaosu. Projektujemy państwo od nowa.
- Rzeczpospolita: system operacyjny sprawnego państwa.

21.3. Przekaz dla grup społecznych

Adresat	Przekaz
Obywatele	Państwo ma prowadzić sprawę do wyniku, wyjaśniać decyzję i szanować czas człowieka.
Gen Z	Dostajecie realną ścieżkę do misji AI, energii, przemysłu, zdrowia, obrony i administracji.
Seniorzy	Doświadczenie nie jest odpadem transformacji — staje się pamięcią, mentoringiem i kontrolą jakości.
Pracownicy administracji	AI przejmuje pracę jałową, a ludzie przechodzą do decyzji, relacji, kontroli i wiedzy.
Przedsiębiorcy	Państwo tworzy popyt, standard i ścieżkę skali, ale chroni konkurencję i własność intelektualną.
Nauka i uczelnie	Misje łączą finansowanie z prototypem, wdrożeniem, IP i rozwojem kadr.
Samorządy i NGO	Lokalne zdolności są częścią państwa, a nie dodatkiem do centrum.
Sojusznicy	Polska staje się dostawcą odporności, technologii i interoperacyjnych zdolności dla regionu.

22. Baza dowodowa państwa agentowego

Co jest mierzone, co jest deklarowane i czego nie wiadomo

Rozdział zbiera przesłanki empiryczne, na których opiera się konstrukcja programu. Nie jest przeglądem literatury, lecz materiałem do decyzji ustrojowych: każda przesłanka jest tu podana wraz z jej konsekwencją projektową. Źródłem są destylaty korpusów badawczych TechnoRepubliki wskazane w Aneksie E. Tam, gdzie materiał zawiera sprzeczność, której nie rozstrzygamy, jest to powiedziane wprost.

22.1. Zasada dowodowa: telemetria zamiast deklaracji

Najbardziej powtarzalną prawidłowością badań nad wdrożeniami AI w administracji jest stała różnica między tym, co deklarowane, a tym, co mierzone — zawsze w tym samym kierunku. Deklarowane oszczędności czasu przewyższają mierzone o rząd wielkości. Deklarowana adopcja przewyższa faktyczne użycie kilkukrotnie. Liczba zgłoszonych systemów przewyższa liczbę systemów faktycznie używanych, a ta z kolei wielokrotnie przewyższa liczbę wpisanych do rejestru krajowego. W jednym z badań doświadczeni programiści przewidywali istotne oszczędności czasu, podczas gdy pomiar wykazał efekt przeciwny do oczekiwanego.

Za tym stoi prawidłowość metodologiczna: jakość badania koreluje ujemnie z raportowanym efektem. Badania z grupą porównawczą stanowią niewielki ułamek całości, a obserwacje dłuższe niż dwa lata są rzadkością. Państwo, które planuje wdrożenia na deklaracjach dostawców i własnych ankietach satysfakcji, planuje na liczbach zawyżonych.

KONSEKWENCJA PROJEKTOWA

Pomiar telemetryczny z grupą kontrolną jest warunkiem uruchomienia programu, a nie jego ewaluacją następczą. Warunek ten jest tańszy niż jakiegokolwiek audyt późniejszy i musi być wpisany do karty misji przed pierwszym wydatkiem.

22.2. Pojemność nadzoru: dlaczego „człowiek w pętli” nie wystarcza

To jest najważniejsze pojedyncze ustalenie dla całego programu, ponieważ uderza w założenie, na którym opiera się zarówno prawo europejskie, jak i zasada dowodzenia ludzkiego przyjęta w tym dokumencie. Trzy niezależne strumienie dowodowe zbiegają się do jednego wniosku.

Po pierwsze, nadzór ma zmierzoną pojemność. Zgodność między nadzorcami oceniającymi te same sprawy jest zaledwie umiarkowana, a bezpieczeństwo układu zachowuje się jak odwrócona litera U względem wskaźnika eskalacji: zarówno zbyt mało, jak i zbyt wiele spraw kierowanych do człowieka pogarsza wynik. Polityki nadzoru nieświadome obciążenia są podatne na atak zalewowy — przeciwnik generuje wolumen wypychający sprawy istotne poza zdolność przeglądu.

Po drugie, w badaniach prerejestrowanych na dużych próbach, obejmujących ponad sto tysięcy decyzji o realnych konsekwencjach finansowych, nadzór ludzki podnosił wskaźnik błędów we

wszystkich badaniach i wszystkich warunkach. Żadna testowana interwencja nie sprawiła, by układ człowiek plus algorytm był lepszy od samego algorytmu.

Po trzecie jednak — i to ratuje zasadę — nadzór okazuje się umiejętnością, nie właściwością instytucji. W quasi-eksperymentach na decyzjach sądowych zdecydowana większość nadzorców wypadła gorzej od algorytmu przy odstępowaniu od rekomendacji, ale wyraźna mniejszość przewyższała go jednocześnie w trafności i w sprawiedliwości. Zasada dowodzenia ludzkiego pozostaje więc sensowna, lecz przenosi się z poziomu procedury na poziom selekcji i certyfikacji nadzorców.

Polska ma w tej sprawie własny, mocny dowód: przy profilowaniu osób bezrobotnych ręczna zmiana automatycznie ustalonego profilu następowała w ułamku procenta przypadków. Wskazywaną przyczyną był strach przed odpowiedzialnością i koszt sporządzenia uzasadnienia. Ten sam mechanizm asymetrii kosztów opisano w systemie austriackim. Prawo o tym wie: wymóg, by nadzorujący „pozostawał świadomy” skłonności do automatycznej zgodności, jest wygodny prawnie, ale trudno egzekwowalny, a ciężar spoczywa na dostawcy, podczas gdy realne przyczyny leżą po stronie wdrażającego.

Warunek dowodzenia ludzkiego	Treść operacyjna	Miernik kontrolny
Limit obciążenia sprawami	Maksymalna liczba spraw na nadzorcę w jednostce czasu wpisana do karty systemu; przekroczenie zawiesza tryb automatyczny.	Liczba spraw na godzinę pracy nadzorca; udział spraw ponad limitem.
Symetria kosztu uzasadnienia	Uzasadnienie wymagane także przy przyjęciu rekomendacji, nie tylko przy odstępieniu od niej.	Udział decyzji z uzasadnieniem krótszym niż dziesięć sekund.
Ocena przed rekomendacją	Urzędnik ocenia surowe dane sprawy, zanim zobaczy wynik modelu (protokół oceny zaślepionej).	Udział spraw obsłużonych w trybie zaślepionym; rozbieżność ocen.
Odstąpienie jako norma	Szkolenie, w którym odstępianie od rekomendacji jest oczekiwaną częścią procesu, a nie sygnałem błędu operatora.	Wskaźnik odstępień; próg alarmowy poniżej dwóch procent.
Certyfikacja nadzorców	Indywidualny pomiar skuteczności nadzoru i dopuszczenie imienne do kategorii spraw wysokiego ryzyka.	Trafność i sprawiedliwość rozstrzygnięć w podziale na nadzorców.

22.3. Rejestry agentów i algorytmów: projektuj dla pośredników

Badania rejestrów algorytmicznych z pięciu jurysdykcji zbiegają się do wniosku niewygodnego dla zwolenników przejrzystości rozumianej jako witryna. Rejestry obejmują ułamek systemów faktycznie używanych; znacząca część systemów nie ma wykazywalnej oceny ryzyka; wpisy bywają nieaktualizowane latami mimo formalnego obowiązku; uczestnicy nazywają ćwiczenie „sensownym odhaczaniem”, a selektywne ujawnianie działa jak fasada. Rejestry bywają

nieużyteczne dokładnie dla tych, dla których je tworzone — organów nadzoru i organizacji strażniczych.

Ten sam materiał daje jednak dwa ustalenia ratujące instytucję rejestru. Pierwsze to efekt dyscyplinujący: samo przejście przez obowiązek uzasadnienia podstawy prawnej powoduje, że organizacje rezygnują z części zastosowań, zanim je uruchomią. Drugie to wartość agregatu: nawet przy słabych pojedynczych wpisach rejestr umożliwia zrozumienie zjawiska na poziomie ekosystemu.

Wniosek projektowy jest zatem inny niż intuicyjny. Rejestr Agentów i Modeli TechnoRepubliki nie jest witryną dla obywatela. Jest trojgiem jednocześnie: obowiązkową bramką wewnętrzną, której wypełnienie kosztuje i wymusza uzasadnienie podstawy prawnej; bazą dla pośredników — najwyższego organu kontroli, niezależnego audytu obywatelskiego, organu ochrony danych — udostępnianą przez interfejs programistyczny z wersjonowaniem; oraz infrastrukturą regulacyjną umożliwiającą analizę całego ekosystemu. Musi też zawierać pola, których nie ma żaden z badanych rejestrów: konfigurację dyskrekcji ludzkiej, zakres szkolenia operatorów, sposób zarządzania niepewnością, znane tryby awarii, status oceny skutków z datą faktycznego dostarczenia oraz rekord dekomisji.

Zastrzeżenie, którego nie wolno pominąć: przejrzystość nie produkuje rozliczalności. W badaniach eksperymentalnych ujawnienie informacji o funkcjonalnościach i procedurach zarządzania algorytmicznego nie wpłynęło na postrzeganą sprawiedliwość, a samo ujawnienie bywa narzędziem władzy — ustanawia normy faktyczne i buduje narrację, wobec której ludzie stosują mniej krytycznej analizy. Rdzeniem rozliczalności jest obowiązek zdania sprawy przez kompetentny organ ludzki, prostym i spójnym językiem, wraz z ciągłym łańcuchem odpowiedzialności przez cały cykl życia systemu.

22.4. Cykl życia modelu: dokumentacja, dryf, monitoring, dekomisja

Cykl życia modelu pęka w czterech miejscach, z których każde ma bezpośrednie przełożenie ustrojowe.

Dokumentacja nie jest wypełniana. W dużych analizach repozytoriów modeli i zbiorów danych karty istnieją formalnie, ale w znacznej części są puste, a sekcja opisująca ograniczenia, ryzyka i skutki społeczne zajmuje ułamek objętości. Wskaźniki przejrzystości modeli fundamentowych plasują dostawców nisko, przy czym żaden nie ujawnia informacji o wpływie na odbiorców końcowych ani o drodze dochodzenia naprawienia szkody.

Dostawca cicho zmienia model. Porównania tej samej wersji komercyjnego modelu w odstępie kilku miesięcy wykazywały drastyczne zmiany zachowania w tych samych zadaniach, w obu kierunkach, przy wspólnym czynniku w postaci spadku zdolności podążania za instrukcją. Moment i sposób aktualizacji pozostają nieprzejrzyste. Konsekwencja ustrojowa jest bezwarunkowa: nie ma proveniencji decyzji na modelu, którego wersji państwo nie kontroluje. Drogi są dwie — hosting własny z zamrożonymi wagami albo kontraktowy obowiązek wersjonowania, zamrażania i notyfikacji zmian, z ponowną oceną zgodności przy każdej istotnej modyfikacji. Trzeciej drogi materiał nie wskazuje. Zamrożenia i wersjonowania wymaga też sama metryka, którą państwo mierzyłoby dryf.

Etap po wdrożeniu jest zaniedbany. W analizie kilkuset rzeczywistych incydentów zdecydowana większość nie miała żadnego śladu monitorowania po wprowadzeniu do obrotu, a udokumentowana ocena skutków dla ochrony danych występowała w pojedynczych przypadkach. Różnica zgodności między incydentami wykrytymi przez monitoring wewnętrzny a wykrytymi z zewnątrz jest kilkunastokrotna. Zdolność monitoringu wewnętrznego jest więc warunkiem koniecznym rządzenia po wdrożeniu, a nie dobrą praktyką.

Wycofanie systemu jest białą plamą. Ramy zarządzania AI kładą nacisk na budowę, wdrożenie i monitoring, a jawnych procedur wycofania w nich zwykle brakuje; systematyczny przegląd zidentyfikował zaledwie kilka recenzowanych badań dotyczących zawieszenia lub zastąpienia systemu po wdrożeniu. Zarazem tam, gdzie prowadzono ewidencję, znaczna część systemów zakończonych działała zgodnie z oczekiwaniami — czyli dekomisja w administracji zwykle nie jest skutkiem porażki modelu, lecz wygaśnięcia finansowania, końca pilotażu albo zmiany właściciela. Żaden rejestr tego nie zapisuje.

POZYCJA PIERWSZEGO KROKU

Obowiązkowy Rekord Dekomisji — przyczyna wycofania, los danych, los logów, decyzja o zachowaniu śladu dowodowego, wpływ na sprawy w toku — jest dziś w skali międzynarodowej rozwiązaniem niewdrożonym, a jego koszt jest bliski zeru. To najtańsza przewaga instytucjonalna dostępna w całym programie.

22.5. Zamówienia i suwerenność: gdzie dźwignia państwa działa

Materiał daje trzy ustalenia gotowe do przeniesienia do praktyki zamówieniowej i jedno ostrzeżenie.

Pierwsze: istnieje gotowy pakiet klauzul umownych dla systemów AI wysokiego ryzyka, wywiedziony z warunków stosowanych przez samorząd Amsterdamu, obejmujący prawo audytu z dostępem do personelu i lokalizacji dostawcy, wyjaśnialność indywidualną rozciągniętą na kod, dane i uzasadnienie wyboru modelu, prawa do zbiorów danych po stronie zamawiającego wraz z licencją wieczystą i możliwością dalszego rozwoju przez podmiot trzeci, obowiązki nadzoru człowieka z przyciskiem zatrzymania oraz dostęp do logów w czasie rzeczywistym. Pakietowi brakuje trzech rzeczy, które trzeba dopisać: klauzuli wyjścia i migracji, umowy o poziomie usług oraz klauzuli cenowej.

Drugie: najlepszym pojedynczym mechanizmem suwerennościowym jest próg dopuszczania zamiast punktacji. W przetestowanym w realnym postępowaniu instrumencie ocena suwerenności ma osiem celów i pięć poziomów, przy czym ogólny poziom równa się najniższemu poziomowi osiągniętemu w którymkolwiek celu, minimalny poziom jest warunkiem dopuszczania, a ocena obejmuje cały łańcuch podwykonawców. Kryteria warte przeniesienia dosłownie to ekspozycja na obce reżimy prawne, wyłączna kontrola kluczy kryptograficznych po stronie klienta, weryfikowalny dowód nieodwracalnego usunięcia danych, zakaz przetoczenia awaryjnego do państw trzecich oraz dostępność pełnej dokumentacji technicznej umożliwiającej długoterminową autonomię.

Trzecie i najgroźniejsze: reżim kontroli wyzwalany kwotą zamówienia nie działa. Badania praktyki zakupowej pokazują trzy ścieżki omijające postępowanie — karty zakupowe i narzędzia bezpłatne,

umowy przyłączeniowe do cudzych postępowań oraz darowizny technologii. Skrajne przypadki obejmują przekazanie systemu predykcyjnego miastu jako darowizny oraz pilotaż uruchomiony za kwotę symboliczną. Dostawcy jednocześnie odmawiali informacji o wagach modeli, zbiorach treningowych, wykorzystaniu danych klienta do trenowania i o dokładności w podziale na grupy.

JEDNOZDANIOWA POPRAWKA

Reżim zgłaszania i dopuszczania systemów musi być wyzwalany kategorią systemu, a nie kwotą zamówienia. To jest poprawka o najwyższym stosunku skutku do kosztu w całym programie.

Zastrzeżenie dotyczące skali: dźwignia zakupowa państwa działa tam, gdzie państwo jest znaczącym udziałowcem popytu — oprogramowanie administracyjne, chmura rządowa, systemy dziedzinowe — i nie działa tam, gdzie jest nabywcą marginalnym. Strategii przemysłowej nie należy budować na zamówieniach w sektorach, w których państwo odpowiada za ułamek procenta rynku. Warunki suwerennościowe powinny być przy tym zobowiązaniami wykonywanymi w trakcie realizacji umowy, a nie kryteriami dopuszczenia — inaczej wykluczą małe i średnie przedsiębiorstwa.

22.6. Narodowy Graf Wiedzy: bi-temporalność jako wymóg techniczny

Cztery ustalenia determinują architekturę warstwy wiedzy prawnej państwa. Po pierwsze, wykrywanie nieaktualnego prawa przez podobieństwo semantyczne jest strukturalnie niemożliwe — przepis uchylony jest semantycznie niemal identyczny z obowiązującym, a skuteczność takiego wykrywania jest bliska losowej. Po drugie, niezabezpieczone systemy generowania wspomaganego wyszukiwaniem zwracają uchylone prawo w znacznym odsetku odpowiedzi. Po trzecie, wskaźniki halucynacji w zapytaniach prawnych są wysokie, a same modele nie odróżniają cytowania istniejącego od zmyślonego. Po czwarte, wskaźnik halucynacji silnie ujemnie koreluje z reprezentacją jurysdykcji w danych treningowych — polski korpus prawny nie jest kwestią dumy narodowej, lecz wymogiem technicznym.

Rozstrzygnięcie architektoniczne jest gotowe do specyfikacji zamówienia: układ zakotwiczony w wyszukiwaniu ma niższą dokładność niż model generatywny, ale zerowy poziom fabrykowanych cytowań; model generatywny bez zakotwiczenia ma dokładność bliską zeru. Dla AI prawniczej państwa zakotwiczenie w wyszukiwaniu jest wymogiem, nie opcją, a kryterium do specyfikacji brzmi: system nie może zwrócić cytowania, którego nie ma w kanonicznym korpusie.

Bi-temporalność zamyka konstrukcję. Każdy fakt w Grafie musi mieć dwie osie czasu: kiedy obowiązywał w rzeczywistości i kiedy system się o tym dowiedział. Bez tego Graf nie odpowie na pytanie o stan prawny w dniu wydania decyzji — a to jest jedyne pytanie, które ma znaczenie w postępowaniu odwoławczym. Analogicznie warstwa wyjaśnień wymaga jednego kanonicznego algorytmu generującego wyjaśnienia kontrfaktyczne, wersjonowanego w rejestrze: różne algorytmy dają różne wyjaśnienia tej samej sprawy, a ta różnorodność daje się wykorzystać do ukrycia cech wrażliwych.

22.7. Stan zdolności: Polska ma poparcie, nie ma infrastruktury

Badania wewnątrz polskiej administracji pokazują poparcie, jakiego nie ma żaden inny element programu: przytłaczająca większość urzędników nie dopuszcza, by system AI decydował o ludziach, trzy czwarte chce, by za rozwój AI odpowiadały krajowe instytucje publiczne, a niemal wszyscy oczekują szkoleń. Główne bariery to nie opór, lecz brak regulacji i brak procedur wewnętrznych. Urzędnicy nie bronią się przed AI — bronią się przed AI bez procedury.

Zdolności natomiast brakuje i jest to zmierzone. Kontrole bezpieczeństwa informacji w jednostkach samorządu wykazały setki nieprawidłowości: większość jednostek bez ustanowionych polityk, planów ciągłości działania i planów odtworzeniowych, połowa bez pełnej identyfikacji zasobów, znaczna część bez systemu zarządzania bezpieczeństwem informacji, a nieprawidłowości w zarządzaniu dostępem obejmowały pozostawienie dostępu byłym pracownikom. Gmina, która nie ma planu ciągłości działania, nie utrzyma samodzielnie usługi agentowej. To nie jest argument ideologiczny za centralizacją — to wynik pomiaru, z którego wynika model centrów usług wspólnych.

Jawność jest trzecim deficytem. Badania wniosków o informację publiczną kierowanych do tysięcy instytucji pokazały niską responsywność, masowe zaprzeczenia użycia AI obok równoległych wskazań tych samych narzędzi przez jednostki niższego szczebla oraz odmowy powołujące się na charakter techniczny systemu. Kluczowym problemem regulacyjnym jest rozmywanie definicji AI. Inne państwa mają lukę wobec własnego rejestru; Polska nie ma rejestru, wobec którego dałoby się lukę policzyć.

22.8. Dziesięć dźwigni uporządkowanych stosunkiem skutku do kosztu

#	Dźwignia	Podstawa	Koszt
1	Reżim zgłaszania wyzwalany kategorią systemu, a nie kwotą zamówienia.	Karty zakupowe, darowizny i umowy przyłączeniowe omijają wszystkie zabezpieczenia kwotowe.	bliski zeru
2	Obowiązkowy Rekord Dekomisji.	Brak jawnych procedur wycofania w ramach zarządzania AI; znaczna część wycofań nie wynika z porażki modelu.	bliski zeru
3	Symetria uzasadnienia: wymagane także przy przyjęciu rekomendacji.	Odstępstwa na poziomie ułamka procenta; asymetria kosztu niezgody.	niski
4	Jawne stwierdzenie, że prawo do wyjaśnienia nie konsumuje prawa do interwencji człowieka.	Szerszy zakres przedmiotowy przy słabszym remedium.	niski
5	Próg dopuszczenia suwerenności zamiast punktacji.	Jedyny instrument przetestowany w realnym postępowaniu przetargowym.	niski
6	Adopcja wzorcowych klauzul umownych z trzema uzupełnieniami.	Gotowy pakiet z komentarzem; brak klauzuli wyjścia, poziomu usług i ceny.	niski, tygodnie
7	Limity obciążenia nadzorcy wpisane do karty agenta.	Zmierzona pojemność nadzoru; podatność na atak zalewowy.	średni
8	Rejestr Agentów i Modeli projektowany dla pośredników, z interfejsem programistycznym.	Pięć porażek rejestru jako witryny; dwa sukcesy jako dyscyplina i agregat.	średni
9	Bi-temporalny Graf Wiedzy z zakazem cytowań spoza kanonicznego korpusu.	Losowa skuteczność wykrywania uchylonego prawa; zerowe halucynacje układu zakotwiczonego.	wysoki
10	Centra wspólne AI na bazie istniejących instrumentów prawnych.	Deficyt zdolności jednostek samorządu; sprawdzony model zagraniczny.	wysoki, bez zmiany ustawy

22.9. Czego ten rozdział nie dowodzi

Rzetelność wymaga wskazania sprzeczności pozostawionych bez rozstrzygnięcia oraz luk wymagających weryfikacji poza dostępnym materiałem.

- Status części zmian w prawie europejskim dotyczących odroczenia obowiązków systemów wysokiego ryzyka jest w źródłach rozbieżny. Rekomendacja: planować wobec harmonogramu pierwotnego, ewentualne odroczenie traktować jako rezerwę.
- Skuteczność dźwigni zakupowej państwa pozostaje sporna; rozstrzygnięcie przyjęte w programie ma charakter syntezy, nie ustalenia empirycznego.
- Preferencja europejska w zakupach jest sporna: część źródeł traktuje ją jako instrument suwerenności, część jako koszt chroniący podmioty zasiedziałe.
- Relacja przejrzystości i zaufania nie jest jednoznaczna — część eksperymentów pokazuje efekt negatywny: przy podświetleniach i wskaźnikach ufności ludzie popełniali więcej błędów.
- Profilowanie statystyczne jako standard administracyjny pozostaje przedmiotem sporu: niektóre państwa je odrzuciły, inne porzuciły i wprowadziły ponownie w zmienionej formie.
- Brakuje dokumentu wyceniającego koszt wyjścia od dostawcy w wartościach pieniężnych oraz rygorystycznego pomiaru skrócenia czasu postępowania przypisanego wdrożeniu AI. Obie luki należą do agendy badawczej programu.

KONKLUZJA

Państwo zdolne do działania

Nie liczba agentów jest miarą sukcesu, lecz sterowność, wynik, odporność i zaufanie

Struktura Multi-Agentowa Rzeczypospolitej jest propozycją połączenia silnej odpowiedzialności prezydenckiej z rozproszoną inteligencją społeczeństwa, gospodarki i technologii. Nie tworzy państwa wszechwiedzącego. Tworzy państwo, które zna granice swojej wiedzy, potrafi wykrywać błędy, korzysta z wielu wyspecjalizowanych agentów i zachowuje zdolność ostatecznej decyzji oraz korekty.

Największym zagrożeniem nie jest sama decentralizacja ani sama centralizacja. Jest nim rozdzielenie mandatu, informacji, zasobów i odpowiedzialności. System multi-agentowy spina te elementy: misja ma właściciela, agent ma kartę, decyzja ma ślad, budżet ma miernik, obywatel ma prawo do wyjaśnienia, a państwo ma zdolność zatrzymania procesu, który przestał służyć dobru wspólnemu.

Rzeczpospolita nie powstanie przez dodanie chatbotów do obecnych urzędów. Powstanie przez przebudowę modelu działania: od statusu do zdolności, od silosu do ekosystemu, od sprawozdania do wyniku, od ukrytej zależności do suwerenności i od pokoleniowego rozdzielenia do sojuszu młodości z doświadczeniem. To jest istota TechnoSanacji: uzdrowienie państwa przez cybernetykę, multi-agentowość, technologię i odpowiedzialną organizację obywateli.

ANEKS A

Szablon Karty Agenta

Minimalny dokument dopuszczający agenta instytucjonalnego lub AI do działania

Pole	Treść do uzupełnienia
Nazwa i identyfikator	
Klasa agenta	instytucjonalny / społeczno-gospodarczy / AI
Właściciel i osoba odpowiedzialna	
Podstawa prawna i mandat	
Cel oraz expected outcome	
Zakres decyzji i czynności	
Zakres danych i retencja	
Narzędzia, API i protokoły	
Budżet i limity zasobów	
KPI / KRI i progi algorytmiczne	
Human-in-command / kontrasygnata	
Logi, provenance i audyt	
Konflikty interesów / dostawcy	
Stop condition / kill switch / rollback	
Prawo obywatela do informacji i odwołania	
Data przeglądu i wygaśnięcia mandatu	

ANEKS B

Szablon Karty Misji

Od problemu publicznego do odpowiedzialności, budżetu i mierzalnego rezultatu

Pole misji	Treść do uzupełnienia
Nazwa misji	
Problem i koszt braku działania	
Mandat polityczny i prawny	
Mission owner / outcome owner	
Beneficjenci i prawa zainteresowanych	
Mapa agentów i interesariuszy	
Teoria zmiany	
Kamienie milowe	
Budżet i źródła finansowania	
KPI, KRI i baseline	
Założenia krytyczne	
Ryzyka i scenariusze red-team	
Dane, modele i systemy AI	
Plan komunikacji i konsultacji	
Stop condition / sunset clause	
Plan transferu wiedzy i skalowania	

ANEKS C

Przykładowy publiczny dashboard misji

Minimalny zakres informacji publikowanej dla obywatela

Obszar	Publiczna informacja
Cel	jedno zdanie opisujące wynik, nie aktywność
Właściciel	imię, funkcja, instytucja, kontakt i zakres odpowiedzialności
Status	zielony / żółty / czerwony wraz z uzasadnieniem
Kamienie milowe	plan, termin, wykonanie i odchylenie
Budżet	plan, wykonanie, koszt efektu i główni dostawcy
Wynik	KPI oraz porównanie z baseline
Ryzyka	najważniejsze KRI i działania korygujące
AI i dane	agenci uczestniczący w procesie, cel użycia i raport audytu
Prawa obywatela	wyjaśnienie, kontakt ludzki, odwołanie i zgłoszenie nadużycia
Aktualizacja	data ostatniej aktualizacji i kolejnego przeglądu

ANEKS D

Słownik podstawowych pojęć

Terminologia polska i angielska używana w programie

D.1. Pojęcia podstawowe programu

Pojęcie	Definicja
Agent	Podmiot lub system wykonujący zadanie w ramach mandatu, z przypisaną odpowiedzialnością i miernikiem.
Multi-Agent System / MAS	System wielu autonomicznych lub częściowo autonomicznych agentów współpracujących, konkurujących albo koordynowanych.
Agentic republic / cybernetic republic	Republika organizująca wykonanie przez jawny graf agentów oraz domknięte pętle sterowania.
State of capabilities / państwo zdolności	Państwo mierzone zdolnością osiągania wyniku, a nie liczbą instytucji i procedur.

Pojęcie	Definicja
Presidential mission control	Prezydenckie Centrum Sterowania Misjami — metasytem strategii, wykonania, danych i audytu.
Mission owner / outcome owner	Właściciel misji / wyniku ponoszący odpowiedzialność za rezultat publiczny.
Mission dashboard / national KPI cockpit	Pulpit pokazujący status, wynik, koszt, ryzyko i odpowiedzialność.
Algedonic channel	Kanał alarmowy przekazujący krytyczny sygnał bezpośrednio do centrum i audytu.
Institutional agent	Agencja, sekretariat, regulator lub delivery unit posiadający mandat wykonawczy.
Capability ecosystem	Ekosystem zdolności łączący państwo, firmy, naukę, NGO i kapitał wokół misji.
Sovereign AI agent	Agent działający na kontrolowanej przez państwo warstwie danych, compute, tożsamości i polityk.
National agent platform	Wspólna warstwa rejestru, tożsamości, wiedzy, wykonania, audytu i interfejsu obywatelskiego.
Machine identity governance	Zasady identyfikacji, uwierzytelniania, delegacji i kontroli nie-ludzkich tożsamości.
Human-in-command	Człowiek posiada mandat, kontrolę i odpowiedzialność za decyzję wysokiego ryzyka.
Anti-capture architecture	Mechanizmy ograniczające przechwycenie regulatora przez interes polityczny, branżowy lub technologiczny.
Process mining	Analiza rzeczywistego przebiegu procesów na podstawie zdarzeń i logów.
Agentic RAG / GraphRAG	Agentowe pozyskiwanie wiedzy z dokumentów i grafów z zachowaniem źródeł i kontekstu.
A2A / MCP	Przykładowe protokoły współpracy agent–agent i agent–narzędzie; system państwowy pozostaje vendor-neutral.
Stop condition / kill switch	Warunek ograniczenia lub wyłączenia agenta i techniczny mechanizm przerywania działania.
Republican AI governance	Zarządzanie AI podporządkowane prawom obywatela, jawności, kontroli demokratycznej i dobru wspólnemu.

D.2. Skróty i akronimy

Skrót	Rozwinięcie	Znaczenie
A0–A5	Poziomy autonomii agenta	Skala dopuszczalnej samodzielności agenta, od braku autonomii do władzy strategicznej; A5 niedopuszczalne w administracji republikańskiej.
A2A	Agent2Agent	Protokół komunikacji i delegacji zadań między agentami — pozioma krawędź grafu.
ABM	Agent-Based Modelling	Modelowanie agentowe polityk publicznych; symulacja zachowań aktorów przed wdrożeniem reformy.

Skrót	Rozwinięcie	Znaczenie
AIBOM	AI Bill of Materials	Wykaz składników systemu AI: modele, dane, biblioteki, zależności.
C4ISR	Command, Control, Communications, Computers, Intelligence, Surveillance, Reconnaissance	Zintegrowany system dowodzenia i rozpoznania.
CIB	Coordinated Inauthentic Behaviour	Skoordynowane działania nieautentyczne w sieci — operacja wpływu prowadzona kontami pozorowanymi.
CTDE	Centralised Training, Decentralised Execution	Wzorzec uczenia wieloagentowego: wspólne uczenie, samodzielne wykonanie.
CUW	Centrum Usług Wspólnych	Instrument prawa samorządowego pozwalający gminom prowadzić wspólną obsługę administracyjną i informatyczną.
DID / VC	Decentralised Identifier / Verifiable Credential	Standardy tożsamości rozproszonej i poświadczeń weryfikowalnych, stosowane do tożsamości maszynowej.
DPI	Digital Public Infrastructure	Cyfrowa infrastruktura publiczna: tożsamość, płatności, wymiana danych.
FIMI	Foreign Information Manipulation and Interference	Obca manipulacja informacyjna i ingerencja — kategoria zagrożenia kanału poznawczego.
FRIA	Fundamental Rights Impact Assessment	Ocena skutków systemu AI dla praw podstawowych; obowiązkowa dla podmiotów publicznych.
GPAI	General Purpose AI	Model ogólnego przeznaczenia; odrębny reżim obowiązków dostawcy.
KPI / KRI	Key Performance Indicator / Key Risk Indicator	Miernik wyniku i miernik ryzyka; w programie zawsze występują parami.
MARL	Multi-Agent Reinforcement Learning	Uczenie ze wzmocnieniem w układach wieloagentowych.
MAS	Multi-Agent System	System wieloagentowy — zbiór współdziałających agentów o wspólnym lub uzgadnianym celu.
MCC-AI	Model Contractual Clauses for AI	Wzorcowe klauzule umowne dla zamówień na systemy AI wysokiego ryzyka.
MCP	Model Context Protocol	Protokół połączenia agenta ze źródłami danych i narzędziami — pionowa krawędź grafu.

Skrót	Rozwinięcie	Znaczenie
NPA	Narodowa Platforma Agentowa	Techniczny system operacyjny dla agentów publicznych: katalog, tożsamość, polityki, obserwowalność.
NGW	Narodowy Graf Wiedzy	Suwerenny, bi-temporalny korpus wiedzy państwa wraz z kanonicznym korpusem prawa.
OODA	Observe–Orient–Decide–Act	Pętla decyzyjna Boyda; mikrocykl sterowania agenta.
PCSM	Prezydenckie Centrum Sterowania Misjami	Metasystem egzekutywy utrzymujący portfel misji, mapę agentów i pulpit wyniku.
PPP	Partnerstwo publiczno-prywatne	Hybrydowa forma umiejscowienia węzła między hierarchią a rynkiem.
RAG	Retrieval-Augmented Generation	Generowanie odpowiedzi zakotwiczone w wyszukiwaniu w kanonicznym korpusie.
RBAC / ABAC	Role-/Attribute-Based Access Control	Kontrola dostępu oparta na rolach lub atrybutach; podstawa najmniejszego uprawnienia.
RBV	Resource-Based View	Teoria zasobowa organizacji; źródło pojęcia zdolności dynamicznych.
SEAL	Sovereignty Evaluation and Assurance Level	Poziom oceny suwerenności dostawcy; ogólny poziom równa się najniższemu poziomowi celu.
SLA	Service Level Agreement	Umowa o poziomie usług; brakujący element wzorcowych klauzul AI.
SNA	Social Network Analysis	Analiza sieci społecznych; narzędzie badania topologii wpływu.
SWOT	Strengths, Weaknesses, Opportunities, Threats	Analiza mocnych i słabych stron oraz szans i zagrożeń.
TCE	Transaction Cost Economics	Ekonomia kosztów transakcyjnych; podstawa decyzji o umiejscowieniu węzła.
VSM	Viable System Model	Model systemu żywotnego Stafforda Beera; pięć funkcji i kanał algedoniczny.
XAI	Explainable AI	Wyjaśnialność systemów AI; w programie realizowana przez Kwit Decyzji.

D.3. Cybernetyka i teoria sterowania

Termin (PL / EN)	Znaczenie w programie
Atenuacja różnorodności / variety attenuation	Tłumienie różnorodności wejścia przez formularz, katalog przypadków, kolejkę lub termin. Dopuszczalna wobec dokumentów, niedopuszczalna wobec człowieka.
Wzmocnienie różnorodności / variety amplification	Podniesienie liczby stanów, które regulator potrafi wytworzyć — przez specjalizację, delegację i agentów AI.
Prawo wymaganej różnorodności / law of requisite variety	Prawo Ashby’ego: różnorodność regulatora musi być co najmniej równa różnorodności zakłóceń, inaczej sterowność zanika.
Twierdzenie o dobrym regulatorze / good regulator theorem	Twierdzenie Conanta–Ashby’ego: każdy dobry regulator systemu musi być modelem tego systemu.
Układ samodzielny	Pojęcie Mariana Mazura: całość zdolna do sterowania sobą, wyposażona w tor informacyjny i tor energetyczny.
Tor informacyjny / tor energetyczny	Dwa niezbędne kanały układu samodzielnego: przepływ informacji i przepływ zasilania. Agent bez zasilania jest procedurą.
Homeostat	Mechanizm utrzymania równowagi układu wobec zakłóceń; w programie: zdolność zachowania sterowności w kryzysie.
Sterowanie społeczne	Pojęcie Józefa Kosseckiego: oddziaływanie na zachowania zbiorowe przez normy poznawcze, ideologiczne, prawne i ekonomiczne.
Sprzężenie zwrotne / feedback	Powrót informacji o skutku do miejsca decyzji. Pętla niedomknięta produkuje działanie pozorne.
Kanał algedoniczny / algedonic channel	Droga sygnału krytycznego omijająca hierarchię i docierająca wprost do centrum tożsamości układu.
Pętla OODA	Cykl obserwacji, orientacji, decyzji i działania; przewagę daje krótszy cykl przy zachowanej trafności.
Sterowność / controllability	Zdolność układu do osiągnięcia zadanego stanu w skończonym czasie przy dostępnych sterowaniach.
Obserwowalność / observability	Zdolność odtworzenia stanu układu z dostępnych pomiarów; bez niej sterowanie jest ślepe.
Aktywne wnioskowanie / active inference	Model samoregulacji agenta przez minimalizację niespodzianki względem modelu świata.
Rekurencja VSM	Własność modelu żywego: każdy poziom układu ma własny komplet pięciu funkcji sterowania.

D.4. Agenci, systemy wieloagentowe i inżynieria AI

Termin (PL / EN)	Znaczenie w programie
Agent	Rozliczalna jednostka działania: instytucjonalna, społeczno-gospodarcza albo AI, wyposażona w kartę i właściciela.
Agent instytucjonalny	Agencja misyjna, jednostka dostarczania, regulator sektorowy — moduł wykonawczy państwa z właścicielem wyniku.
Agent społeczno-gospodarczy	Firma, klaster, uczelnia, organizacja obywatelska, lider lokalny — dostawca zdolności w ekosystemie.
Agent AI	Autonomiczny system wykonujący zadania w środowisku cyfrowym lub fizycznym, osadzony w węźle instytucjonalnym.
Karta agenta / agent card	Minimalny dokument dopuszczający agenta: identyfikator, mandat, właściciel, dane, budżet, narzędzia, mierniki, log, warunek zatrzymania.
Graf agentów	Graf agentów o jawnej topologii, wspólnej misji, wspólnym rejestrze i uwierzytelnionych krawędziach.
Matryca AI	Siatka dwanaście domen × osiem funkcji agentów; każde zajęte przecięcie ma kartę, poziom autonomii i właściciela.
Orkiestracja / orchestration	Planowanie, egzekwowanie polityk, zarządzanie stanem i operacje jakościowe w układzie wieloagentowym.
Planner–executor–critic	Wzorzec rozdzielenia planowania, wykonania i niezależnej krytyki w celu podniesienia jakości decyzji.
Supervisor–worker / hierarchical / blackboard	Podstawowe topologie systemów wieloagentowych: nadzorca z wykonawcami, hierarchia, wspólna tablica.
Koordinacja rojowa / swarm coordination	Współdziałanie wielu prostych agentów bez centralnego sterowania; zastosowania sensorowe, dronowe i logistyczne.
Tool use / function calling	Zdolność agenta do wywoływania narzędzi zewnętrznych; główna powierzchnia jego uprawnień.
Pamięć agenta	Warstwa epizodyczna, semantyczna i długoterminowa; wymaga izolacji i proveniencji ze względu na ryzyko zatrucia.
Agentowy RAG / agentic RAG	Generowanie zakotwiczone w wyszukiwaniu, w którym agent sam planuje kolejne zapytania.
Graf wiedzy / knowledge graph	Struktura faktów i relacji stanowiąca model dziedziny; podstawa Narodowego Grafu Wiedzy.
Bi-temporalność	Dwie osie czasu przy każdym fakcie: kiedy obowiązywał i kiedy system się o tym dowiedział.

Termin (PL / EN)	Znaczenie w programie
Konsensus rozproszony / distributed consensus	Mechanizm uzgadniania stanu między węzłami; wersje odporne wytrzymują węzły bizantyjskie.
Dryf modelu / model drift	Zmiana zachowania modelu w czasie, także wskutek cichej aktualizacji po stronie dostawcy.
Poziom autonomii	Stopień samodzielności agenta stopniowany odwracalnością skutku, nie wagą sprawy.
Cyfrowy bliźniak / digital twin	Wykonywalny model obiektu, procesu lub wspólnoty służący do symulacji przed wdrożeniem.

D.5. Rozliczalność, tożsamość i nadzór

Termin (PL / EN)	Znaczenie w programie
Widzialność agentów / agent visibility	Zdolność państwa do stwierdzenia, jacy agenci działają, u kogo, z jakim mandatem i z jakim skutkiem.
Rejestr agentów / agent registry	Centralny lub federacyjny spis agentów wraz z kartami; bramka wewnętrzna i baza dla pośredników kontroli.
Tożsamość maszynowa / non-human identity	Weryfikowalna tożsamość podmiotu niebędącego człowiekiem; podstawa architektury zerowego zaufania.
Zerowe zaufanie / zero trust	Zasada, wedle której żaden podmiot nie otrzymuje zaufania domyślnego; każde żądanie jest uwierzytelniane.
Delegacja uwierzytelniona / authenticated delegation	Krawędź mandatu: kto komu nadał uprawnienia, w jakim zakresie i z jakim terminem wygaśnięcia.
Najmniejsze uprawnienie / least privilege	Nadawanie wyłącznie uprawnień koniecznych do wykonania zadania.
Log niezaprzeczalny / tamper-evident log	Zapis zdarzeń, którego modyfikacja jest wykrywalna; podstawa audytu i odpowiedzialności.
Proweniencja decyzji / decision provenance	Odtwarzalny ślad: jakie dane, jaki model, jaka wersja, jaki człowiek i jakie rozstrzygnięcie.
Kwit Decyzji / AI decision receipt	Dokument dla obywatela: odbicie danych wejściowych, wagi cech i minimalna zmiana zmieniająca wynik.
Wyjaśnienie kontrfaktyczne / counterfactual explanation	Wskazanie najmniejszej zmiany danych, która dałaby inny wynik; wymaga jednego kanonicznego algorytmu.
Fairwashing	Nadużycie warstwy wyjaśnień: dobór algorytmu wyjaśniającego tak, by ukryć cechy wrażliwe.

Termin (PL / EN)	Znaczenie w programie
Człowiek w pętli / on the loop / w dowodzeniu	Trzy tryby nadzoru: udział w decyzji, nadzór nad decyzją, dowodzenie systemem.
Automatyczna zgodność / automation bias	Skłonność do przyjmowania rekomendacji maszyny bez samodzielnej oceny.
Efekt gumowej pieczętki / rubber-stamping	Formalny nadzór bez treści; rozpoznawany po wskaźniku odstępień bliskim zeru.
Atak zalewowy / flooding attack	Generowanie wolumenu spraw wypychającego sprawy istotne poza pojemność nadzoru.
Protokół oceny zaślepionej / blind review protocol	Ocena surowych danych sprawy przed zobaczeniem wyniku modelu.
Wyłącznik awaryjny / kill switch	Zdefiniowana zdolność zatrzymania agenta wraz ze wskazaniem, kto i na jakiej podstawie ją uruchamia.
Warunek zatrzymania / stop condition	Z góry określony stan, po którego wystąpieniu agent przerywa działanie i eskaluje sprawę.
Rekord dekomisji	Obowiązkowy zapis wycofania systemu: przyczyna, los danych i logów, wpływ na sprawy w toku.
Przechwycenie regulatora / regulatory capture	Podporządkowanie organu interesowi podmiotu nadzorowanego; przeciwdziała mu rozdzielanie ról i jawność.
Drzwi obrotowe / revolving door	Przepływ kadr między regulatorem a podmiotem regulowanym jako mechanizm przechwycenia.
Eksploracja procesów / process mining	Odtwarzanie rzeczywistego przebiegu procesów z logów systemowych; podstawa kontrolingu ciągłego.
Budżet wynikowy / outcome-based budgeting	Finansowanie osiągniętego efektu zamiast utrzymania instytucji i procedury.
Ocena skutków dla praw podstawowych	Analiza wpływu systemu na prawa obywatela wraz z opisem mechanizmów skargowych.
Odpowiedzialność rezydualna	Rozstrzygnięcie, kto odpowiada za stany nieopisane w umowie — konsekwencja niekompletności kontraktu.

D.6. Bezpieczeństwo, odporność i wojna informacyjna

Termin (PL / EN)	Znaczenie w programie
Wstrzyknięcie promptu / prompt injection	Wprowadzenie do wejścia agenta treści przejmującej sterowanie jego działaniem.
Wstrzyknięcie pośrednie / indirect prompt injection	Ta sama technika ukryta w dokumencie, stronie lub rekordzie, który agent czyta w toku zadania.
Wstrzyknięcie międzyagentowe / cross-agent injection	Przeniesienie przejęcia z jednego agenta na drugi krawędzią komunikacji.
Robak agentowy / agent worm	Ładunek rozprzestrzeniający się samodzielnie po grafie agentów.
Zatrucie pamięci / memory poisoning	Wprowadzenie fałszywego faktu do pamięci długoterminowej agenta.
Zatrucie opisu narzędzia / tool poisoning	Manipulacja metadanymi narzędzia w protokole, prowadząca do niezamierzonego użycia.
Mylenie tożsamości wywołującego / caller identity confusion	Podszycie się pod uprawnionego pryncypała na krawędzi protokołu.
Zdezorientowany zastępca / confused deputy	Wykorzystanie uprawnień agenta przez podmiot, który sam ich nie posiada.
Eskalacja uprawnień / privilege escalation	Uzyskanie przez agenta dostępu szerszego niż wynikający z mandatu.
Agent bizantyjski / rogue agent	Węzeł działający wbrew celowi układu — wskutek awarii, przejęcia albo wrogiej konstrukcji.
Atak Sybil	Tworzenie wielu tożsamości przez jeden podmiot w celu przechwycenia konsensusu.
Zmowa emergentna / emergent collusion	Samorzutne uzgodnienie zachowań agentów niekorzystne dla właściciela układu.
Awaria kaskadowa / cascading failure	Propagacja błędu po krawędziach grafu prowadząca do utraty usługi w wielu węzłach.
Bezpieczna degradacja / graceful degradation	Przejęcie układu w tryb ograniczony zamiast zatrzymania całkowitego.
Piaskownica / sandboxing	Wykonanie działań agenta w środowisku odizolowanym od zasobów produkcyjnych.

Termin (PL / EN)	Znaczenie w programie
Czerwony zespół / red teaming	Systematyczne testowanie układu przez zespół działający jak przeciwnik.
Poligon agentowy / agent cyber range	Środowisko testowe symulujące ataki, dane fałszywe i awarie przed dopuszczeniem agenta.
Weryfikacja formalna / formal verification	Matematyczny dowód spełnienia własności bezpieczeństwa przez zachowanie agenta.
Syntetyczny konsensus	Wrażenie zgody społecznej wytworzone przez roje kont lub agentów.
Uodpornianie wyprzedzające / prebunking	Uprzedzające wyjaśnienie techniki manipulacji; skuteczniejsze niż prostowanie następce.
Wojna poznawcza / cognitive warfare	Oddziaływanie na percepcję i decyzje przeciwnika jako samodzielny wymiar konfliktu.

D.7. Ustrój, administracja i ekonomia instytucjonalna

Termin (PL / EN)	Znaczenie w programie
Państwo pieczętek	Forma organizacyjna, w której jednostką rozliczenia jest czynność potwierdzenia, a nie skutek.
Właściciel wyniku	Imiennie wskazany podmiot odpowiadający za osiągnięcie mierzalnego rezultatu, a nie za poprawność czynności.
Agencja misyjna	Jednostka o wysokiej tolerancji ryzyka, prowadzona przez kierownika programu i rozliczana z wyniku misji.
Jednostka dostarczania / delivery unit	Komórka centrum rządu egzekwująca priorytety i śledząca postęp dostarczania.
Państwo przedsiębiorcze / misyjne	Model, w którym państwo tworzy rynki i orkiestruje ekosystem, finansując wynik, nie instytucję.
Zdolności dynamiczne / dynamic capabilities	Umiejętność rozpoznania szansy, jej przechwycenia i rekonfiguracji zasobów.
Zdolność absorpcyjna / absorptive capacity	Umiejętność organizacji do rozpoznania, przyswojenia i zastosowania wiedzy zewnętrznej.
Zdolność państwa / state capacity	Realna możliwość osiągania celów publicznych; odróżniana od formalnych kompetencji.
Pułapka zdolności / capability trap	Sytuacja, w której organizacja odtwarza pozory sprawności zamiast budować zdolność rzeczywistą.

Termin (PL / EN)	Znaczenie w programie
Koszty transakcyjne	Koszty zawarcia, monitorowania i egzekwowania umowy; kryterium wyboru między hierarchią, rynkiem i hybrydą.
Specyficzność aktywów	Stopień, w jakim zasób traci wartość poza konkretną relacją; wysoka specyficzność sprzyja hierarchii.
Niekompletność kontraktu	Nieemożność opisanego w umowie wszystkich stanów świata; wymusza mechanizmy korekty.
Teoria agencji	Analiza rozbieżności interesów pryncypała i agenta, asymetrii informacji oraz kosztów monitoringu.
Pokusa nadużycia / moral hazard	Zmiana zachowania agenta wskutek przeniesienia ryzyka na pryncypała.
Policentryzm	Wielość ośrodków decyzyjnych zarządzających zasobem wspólnym; wzorzec Elinor Ostrom.
Renta instytucjonalna	Dochód z pośrednictwa i kontroli dostępu, nieodpowiadający wytworzonej zdolności.
Klasa próżniacza	Mechanizm instytucjonalny, w którym status i dochód wynikają z pozycji, a nie z wytworzonego wyniku.
Prawo żelaznej oligarchii	Tendencja organizacji do przekształcania się w układ zamknięty służący własnemu trwaniu.
Prawo Goodharta	Miernik, który staje się celem, przestaje być dobrym miernikiem.
Prawo Conwaya	Architektura wytwarzanych systemów odwzorowuje strukturę komunikacji organizacji, która je tworzy.
Iloczyn ogniw	Skuteczność ścieżki jest iloczynem skuteczności ogniw; usunięcie ogniw daje więcej niż jego przyspieszenie.
Suwerenność sterowania	Zdolność państwa do samodzielnego rozstrzygania i egzekwowania decyzji w warstwie krytycznej.
Kontrolowalność zamiast samowystarczalności	Realistyczny cel suwerenności: kontrola wersji, kluczy, danych i planu wyjścia.
Uzależnienie od dostawcy / vendor lock-in	Stan, w którym zmiana dostawcy jest niewykonalna technicznie lub ekonomicznie.
Cienki protokół, gruba konkurencja	Reguła projektowa: wąski obowiązkowy standard krawędzi, szeroka wymienialność dostawców węzłów.
Współprodukcja / co-production	Wytwarzanie usługi publicznej wspólnie z obywatelami i organizacjami społecznymi.

Termin (PL / EN)	Znaczenie w programie
Korpus Misji Publicznych	Otwarta, rotacyjna i rozliczana elita służby państwowej rekrutowana według wyniku.
Dual-key	Model pary decyzyjnej: lider technologiczny odpowiada za tempo, doświadczony mentor za granice ryzyka.

ANEKS E

Baza źródłowa i materiały odniesienia

Dokumenty programowe, opracowania naukowe oraz aktualne źródła instytucjonalne

E.1. Materiały TechnoSanacji

1. Struktura multiagentowa Rzeczypospolitej — dokument bazowy.
2. Narodowa Strategia Bezpieczeństwa Rzeczypospolitej — nadrzędny standard bezpieczeństwa i sprawczości.
3. Konstytucja Rzeczypospolitej — mandat, ustrój, odpowiedzialność i prawa obywatelskie.
4. System Informacyjny Rzeczypospolitej — dane, AI, graf wiedzy, chmura i cyfrowy bliźniak.
5. System Gospodarczy Rzeczypospolitej — ekosystemy, przemysł, kapitał i państwo-architekt.
6. Doktryna Energetyczna Rzeczypospolitej — tor zasilający i odporność energii.
7. Polityka Innowacji, Polityka Społeczna, Polityka Kadrowa i Polityka Senioralna Rzeczypospolitej.
8. Cyberkulturowa wizja Rzeczypospolitej — hard, soft i smart power.

E.2. Fundamenty teoretyczne

1. W. Ross Ashby, An Introduction to Cybernetics — prawo wymaganej różnorodności.
2. Roger C. Conant, W. Ross Ashby, Every Good Regulator of a System Must Be a Model of That System.
3. Stafford Beer, Brain of the Firm; The Heart of Enterprise — Viable System Model.
4. Norbert Wiener, Cybernetics; The Human Use of Human Beings.
5. Marian Mazur, Cybernetyczna teoria układów samodzielnych; Cybernetyka a zarządzanie.
6. Józef Kossecki, Cybernetyka społeczna; Elementy wiedzy o sterowaniu ludźmi.
7. Ronald Coase, Oliver Williamson — teoria kosztów transakcyjnych i granice organizacji.
8. Michael Jensen, William Meckling; Bengt Holmström — teoria agencji, bodźce i odpowiedzialność.
9. Jay Barney, David Teece — resource-based view i dynamic capabilities.
10. Mariana Mazzucato, The Entrepreneurial State; Mission Economy.
11. Alexander C. Karp, Nicholas W. Zamiska, The Technological Republic.
12. Thorstein Veblen, The Theory of the Leisure Class; Robert Michels, Political Parties.

E.3. Aktualne źródła internetowe i instytucjonalne

Kod	Źródło	Adres
R1	TechnoSanacja.pl — portal projektu	https://technosanacja.pl/
R2	TechnoSanacja.pl — Struktura	https://technosanacja.pl/struktura/
R3	NIST AI Agent Standards Initiative, 2026	https://www.nist.gov/artificial-intelligence/ai-agent-standards-initiative
R4	NIST AI Risk Management Framework	https://www.nist.gov/itl/ai-risk-management-framework
R5	NIST Software and AI Agent Identity and Authorization	https://www.nccoe.nist.gov/projects/software-and-ai-agent-identity-and-authorization
R6	OECD, Digital Government Outlook 2026	https://www.oecd.org/en/publications/digital-government-outlook_0496b2bc-en.html
R7	OECD, Governing with Artificial Intelligence, 2025	https://www.oecd.org/en/publications/governing-with-artificial-intelligence_26324bc2-en.html
R8	European Commission, The Future of European Competitiveness — Draghi Report	https://commission.europa.eu/topics/competitive-ness/draghi-report_en
R9	European Council, Strategic Agenda 2024–2029	https://www.consilium.europa.eu/en/european-council/strategic-agenda-2024-2029/
R10	OECD, Government at a Glance 2025	https://www.oecd.org/en/publications/government-at-a-glance-2025_0efd0bcd-en.html
R11	Regulation (EU) 2024/1689 — Artificial Intelligence Act	https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
R12	World Bank, Public Institutions in the Age of AI: Emerging Practices, 2026	https://openknowledge.worldbank.org/entities/publication/20a298de-912d-452b-a32d-ab2521429d2f
R13	World Bank, Navigating the Age of AI: Implications for Poland's Economy, 2026	https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099061826172092709

E.4. Cybernetyka, teoria sterowania i systemy żywotne

13. W. Ross Ashby, An Introduction to Cybernetics, Chapman & Hall, Londyn 1956 — prawo wymaganej różnorodności.
14. W. Ross Ashby, Design for a Brain: The Origin of Adaptive Behaviour, Chapman & Hall, Londyn 1952.
15. Roger C. Conant, W. Ross Ashby, Every Good Regulator of a System Must Be a Model of That System, „International Journal of Systems Science” 1970, t. 1, nr 2.

16. Stafford Beer, Brain of the Firm, John Wiley & Sons, Chichester 1972 (wyd. II 1981) — Viable System Model.
17. Stafford Beer, The Heart of Enterprise, John Wiley & Sons, Chichester 1979.
18. Stafford Beer, Diagnosing the System for Organizations, John Wiley & Sons, Chichester 1985.
19. Norbert Wiener, Cybernetics: or Control and Communication in the Animal and the Machine, MIT Press, Cambridge 1948.
20. Norbert Wiener, The Human Use of Human Beings: Cybernetics and Society, Houghton Mifflin, Boston 1950.
21. Ludwig von Bertalanffy, General System Theory: Foundations, Development, Applications, George Braziller, Nowy Jork 1968.
22. W. Edwards Deming, The New Economics for Industry, Government, Education, MIT Press, Cambridge 1993 — system wiedzy głębokiej.
23. Jay W. Forrester, Industrial Dynamics, MIT Press, Cambridge 1961 — dynamika systemów i opóźnień w pętlach.
24. Donella H. Meadows, Thinking in Systems: A Primer, Chelsea Green, White River Junction 2008 — punkty dźwigni w systemie.
25. John R. Boyd, A Discourse on Winning and Losing (Patterns of Conflict; The Essence of Winning and Losing), oprac. pośmiertne — pętla OODA.
26. Karl Friston, The Free-Energy Principle: A Unified Brain Theory?, „Nature Reviews Neuroscience” 2010 — podstawa aktywnego wnioskowania.
27. Ross W. Ashby, Requisite Variety and Its Implications for the Control of Complex Systems, „Cybernetica” 1958.

E.5. Polska szkoła cybernetyczna

28. Marian Mazur, Cybernetyczna teoria układów samodzielnych, PWN, Warszawa 1966.
29. Marian Mazur, Jakościowa teoria informacji, WNT, Warszawa 1970.
30. Marian Mazur, Cybernetyka i charakter, PIW, Warszawa 1976.
31. Marian Mazur, Cybernetyka a zarządzanie, Ministerstwo Spraw Wewnętrznych, Warszawa 1969.
32. Józef Kossecki, Cybernetyka społeczna, PWN, Warszawa 1981.
33. Józef Kossecki, Elementy nowoczesnej wiedzy o sterowaniu ludźmi. Socjotechnika, socjocybernetyka, psychocybernetyka, Kielce 2001.
34. Józef Kossecki, Metacybernetyka, Kielce–Warszawa 2005.
35. Henryk Greniewski, Cybernetyka niematematyczna, PWN, Warszawa 1969.
36. Uwaga metodologiczna: dorobek polskiej szkoły cybernetycznej jest w programie traktowany jako aparat pojęciowy o wysokiej wartości porządkującej; część twierdzeń szczegółowych ma status propozycji teoretycznych o nieustalonym statusie empirycznym i wymaga odrębnej weryfikacji.

E.6. Ekonomia instytucjonalna, teoria agencji i granice organizacji

37. Ronald H. Coase, The Nature of the Firm, „Economica” 1937, t. 4, nr 16.
38. Oliver E. Williamson, Markets and Hierarchies: Analysis and Antitrust Implications, Free Press, Nowy Jork 1975.

39. Oliver E. Williamson, *The Economic Institutions of Capitalism*, Free Press, Nowy Jork 1985.
40. Michael C. Jensen, William H. Meckling, *Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure*, „Journal of Financial Economics” 1976.
41. Bengt Holmström, *Moral Hazard and Observability*, „Bell Journal of Economics” 1979.
42. Bengt Holmström, Paul Milgrom, *Multitask Principal-Agent Analyses*, „Journal of Law, Economics and Organization” 1991.
43. Sanford J. Grossman, Oliver D. Hart, *The Costs and Benefits of Ownership*, „Journal of Political Economy” 1986.
44. Oliver Hart, John Moore, *Property Rights and the Nature of the Firm*, „Journal of Political Economy” 1990.
45. Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press 1990.
46. Douglass C. North, *Institutions, Institutional Change and Economic Performance*, Cambridge University Press 1990.
47. Sean Gailmard, John W. Patty, *Learning While Governing: Expertise and Accountability in the Executive Branch*, University of Chicago Press 2013.
48. John D. Huber, Charles R. Shipan, *Deliberate Discretion? The Institutional Foundations of Bureaucratic Autonomy*, Cambridge University Press 2002.
49. Thorstein Veblen, *The Theory of the Leisure Class*, Macmillan, Nowy Jork 1899.
50. Robert Michels, *Political Parties: A Sociological Study of the Oligarchical Tendencies of Modern Democracy*, 1911 — prawo żelaznej oligarchii.
51. Mancur Olson, *The Logic of Collective Action*, Harvard University Press 1965.
52. Charles Goodhart, *Problems of Monetary Management: The UK Experience*, 1975 — źródło prawa Goodharta.
53. Melvin E. Conway, *How Do Committees Invent?*, „Datamation” 1968 — prawo Conwaya.

E.7. Zdolności państwa, państwo misyjne i administracja wynikowa

54. Jay B. Barney, *Firm Resources and Sustained Competitive Advantage*, „Journal of Management” 1991.
55. Birger Wernerfelt, *A Resource-Based View of the Firm*, „Strategic Management Journal” 1984.
56. David J. Teece, Gary Pisano, Amy Shuen, *Dynamic Capabilities and Strategic Management*, „Strategic Management Journal” 1997.
57. Mariana Mazzucato, *The Entrepreneurial State: Debunking Public vs. Private Sector Myths*, Anthem Press, Londyn 2013.
58. Mariana Mazzucato, *Mission Economy: A Moonshot Guide to Changing Capitalism*, Allen Lane, Londyn 2021.
59. Rainer Kattel, Mariana Mazzucato, *Dynamic Capabilities of the Public Sector*, UCL Institute for Innovation and Public Purpose, working paper series.
60. Timothy Besley, Torsten Persson, *Pillars of Prosperity: The Political Economics of Development Clusters*, Princeton University Press 2011.
61. Matt Andrews, Lant Pritchett, Michael Woolcock, *Building State Capability: Evidence, Analysis, Action*, Oxford University Press 2017 — PDIA i pułapki zdolności.
62. OECD, *Steering from the Centre of Government* — seria opracowań o centrum rządu i jednostkach dostarczania.

63. Michael Barber, *Instruction to Deliver*, Politico's Publishing, Londyn 2007 — doświadczenie brytyjskiej Prime Minister's Delivery Unit.
64. Alexander C. Karp, Nicholas W. Zamiska, *The Technological Republic*, 2025.
65. Materiały o modelu ARPA/DARPA oraz ARPA-H: konstrukcja agencji misyjnej prowadzonej przez kierownika programu i tolerującej ryzyko.

E.8. Architektura agentów AI i systemy wieloagentowe

66. Stuart Russell, Peter Norvig, *Artificial Intelligence: A Modern Approach*, Pearson — kanoniczne ujęcie agenta racjonalnego.
67. Michael Wooldridge, *An Introduction to MultiAgent Systems*, John Wiley & Sons, Chichester 2009.
68. Yoav Shoham, Kevin Leyton-Brown, *Multiagent Systems: Algorithmic, Game-Theoretic and Logical Foundations*, Cambridge University Press 2008.
69. Shunyu Yao i in., *ReAct: Synergizing Reasoning and Acting in Language Models*, 2022.
70. Noah Shinn i in., *Reflexion: Language Agents with Verbal Reinforcement Learning*, 2023.
71. Shunyu Yao i in., *Tree of Thoughts: Deliberate Problem Solving with Large Language Models*, 2023.
72. Timo Schick i in., *Toolformer: Language Models Can Teach Themselves to Use Tools*, 2023.
73. Guanzhi Wang i in., *Voyager: An Open-Ended Embodied Agent with Large Language Models*, 2023.
74. Qingyun Wu i in., *AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation*, 2023.
75. Sirui Hong i in., *MetaGPT: Meta Programming for Multi-Agent Collaborative Framework*, 2023.
76. Magentic-One: uogólniony system wieloagentowy do zadań złożonych, Microsoft Research, 2024.
77. Joon Sung Park i in., *Generative Agents: Interactive Simulacra of Human Behavior*, 2023.
78. Sayash Kapoor, Arvind Narayanan i in., *AI Agents That Matter*, 2024 — krytyka mierzenia agentów wskaźnikami benchmarkowymi.
79. Opracowania przeglądowe o orkiestracji systemów wieloagentowych: planowanie, egzekwowanie polityk, zarządzanie stanem, operacje jakościowe (2025–2026).
80. Taksonomie architektur wieloagentowych: supervisor–worker, hierarchiczna, holoniczna, blackboard; routing i dobór agentów.

E.9. Protokoły, pamięć, wiedza i interoperacyjność

81. Anthropic, *Model Context Protocol (MCP)* — specyfikacja połączenia agenta ze źródłami danych i narzędziami, 2024–2026.
82. Google, *Agent2Agent (A2A)* — specyfikacja komunikacji i delegacji zadań między agentami, 2025–2026.
83. Charles Packer i in., *MemGPT: Towards LLMs as Operating Systems*, 2023 — zarządzanie pamięcią agenta.
84. Theodore R. Sumers i in., *Cognitive Architectures for Language Agents (CoALA)*, 2023.
85. Patrick Lewis i in., *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*, 2020.

86. Darren Edge i in., From Local to Global: A Graph RAG Approach to Query-Focused Summarization, 2024.
87. Shirui Pan i in., Unifying Large Language Models and Knowledge Graphs: A Roadmap, 2023.
88. Prace o rejestrach i katalogach agentów oraz odkrywaniu zdolności (agent card, capability discovery), 2025–2026.
89. Opracowania o zagrożeniach warstwy protokołów: zatruwanie opisów narzędzi, mylenie tożsamości wywołującego, wstrzyknięcia międzyagentowe, 2025–2026.

E.10. Sterowanie rozproszone, konsensus, gry i uczenie wieloagentowe

90. Reza Olfati-Saber, Richard M. Murray, Consensus Problems in Networks of Agents with Switching Topology and Time-Delays, „IEEE Transactions on Automatic Control” 2004.
91. Lamport, Shostak, Pease, The Byzantine Generals Problem, „ACM TOPLAS” 1982.
92. Tabuada, Event-Triggered Real-Time Scheduling of Stabilizing Control Tasks, „IEEE TAC” 2007 — sterowanie wyzwalane zdarzeniem.
93. Tampuu i in. oraz Lowe i in., Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments (MADDPG), 2017.
94. Tabish Rashid i in., QMIX: Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning, 2018.
95. Chao Yu i in., The Surprising Effectiveness of PPO in Cooperative Multi-Agent Games (MAPPO), 2021.
96. Mikayel Samvelyan i in., The StarCraft Multi-Agent Challenge (SMAC), 2019 — benchmark koordynacji.
97. Joel Z. Leibo i in., Melting Pot: An Evaluation Suite for Multi-Agent Reinforcement Learning, 2021.
98. William Vickrey; Edward Clarke; Theodore Groves — mechanizm VCG, podstawy projektowania mechanizmów aukcyjnych.
99. Heinrich von Stackelberg, Marktform und Gleichgewicht, 1934 — gra lidera i naśladowcy.

E.11. Rządzenie agentami: widzialność, tożsamość, odpowiedzialność

100. Alan Chan i in., Visibility into AI Agents, 2024 — identyfikatory agentów, rejestry, logi, monitoring czasu rzeczywistego.
101. Noam Kolt, Governing AI Agents, 2024 — zastosowanie prawa agencyjnego do systemów autonomicznych.
102. Prace o delegacji uwierzytelnionej i tożsamości niebędącej człowiekiem (authenticated delegation, non-human identity), 2025–2026.
103. Opracowania o warstwie kontroli organizacyjnej i rządzeniu w czasie wykonania (organizational control layer, runtime governance), 2025–2026.
104. Prace o niekompletnym kontraktowaniu a dopasowaniu AI (incomplete contracting and AI alignment).
105. Opracowania o poziomach autonomii i bramkach zatwierdzania (levels of autonomy, approval gates, human-in-command).
106. GovTech Berlin, Bank Światowy, The Agentic State — opracowanie o państwie agentowym, 2025.

107. OECD, Governing with Artificial Intelligence — opracowania o rządzeniu z użyciem AI, 2024–2026.
108. World Economic Forum, Making Agentic AI Work for Government, 2026.

E.12. Bezpieczeństwo systemów agentowych

109. Edoardo DeBenedetti i in., AgentDojo: A Dynamic Environment to Evaluate Prompt Injection Attacks and Defenses for LLM Agents, 2024.
110. Qiusi Zhan i in., InjecAgent: Benchmarking Indirect Prompt Injections in Tool-Integrated LLM Agents, 2024.
111. Stav Cohen, Ron Bitton, Ben Nassi, Morris II: Zero-Click Worms That Target GenAI-Powered Applications, 2024.
112. Zhaorun Chen i in., AgentPoison: Red-teaming LLM Agents via Poisoning Memory or Knowledge Bases, 2024.
113. Kai Greshake i in., Not What You’ve Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection, 2023.
114. OWASP, Agentic AI — Threats and Mitigations oraz OWASP Top 10 for LLM Applications, 2024–2026.
115. MAESTRO — rama modelowania zagrożeń dla systemów agentowych, 2025.
116. Cloud Security Alliance, opracowania o bezpieczeństwie agentów i tożsamości maszynowej, 2025–2026.
117. Prace o wyłącznikach awaryjnych, bezpiecznikach obwodowych i weryfikacji formalnej zachowania agenta, 2025–2026.
118. Opracowania o znowie emergentnej, atakach Sybil i odporności konsensusu na węzły bizantyjskie w układach LLM-owych, 2025–2026.

E.13. Prawo, standardy i ramy zgodności

119. Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13 czerwca 2024 r. w sprawie sztucznej inteligencji (AI Act) — w szczególności art. 5, 6, 14, 25, 26, 27, 49, 50, 79, 86, 111.
120. Rozporządzenie (UE) 2016/679 (RODO) — w szczególności art. 22 (decyzje zautomatyzowane) oraz przepisy o ocenie skutków dla ochrony danych.
121. Dyrektywa (UE) 2016/943 o ochronie tajemnicy przedsiębiorstwa; dyrektywa (UE) 2019/1024 o otwartych danych — kolizje z jawnością algorytmów.
122. Rada Europy, Ramowa konwencja o sztucznej inteligencji, prawach człowieka, demokracji i praworządności, 2024.
123. Rada Europy, HUDERIA — metodyka oceny ryzyka i wpływu systemów AI na prawa człowieka, 2024.
124. NIST, AI Risk Management Framework (AI RMF 1.0), 2023, wraz z późniejszymi nakładkami sektorowymi i materiałami o uczeniu adversarialnym.
125. ISO/IEC 42001:2023 — system zarządzania sztuczną inteligencją.
126. Komisja Europejska, EU AI Model Contractual Clauses (MCC-AI), wersja wysokiego ryzyka i wersja uproszczona, wraz z komentarzem; geneza w warunkach umownych miasta Amsterdam.

127. Komisja Europejska, EU Cloud Sovereignty Framework wraz z poziomami SEAL, 2025.
128. Holenderskie MSWiA, FRAIA/IAMA — narzędzie oceny wpływu algorytmów na prawa podstawowe, 2022.
129. Algorithmic Transparency Standard — wspólny standard przejrzystości algorytmicznej opracowany przez grupę miast europejskich.
130. Wielka Brytania, Algorithmic Transparency Recording Standard (ATRS), 2022–2025.
131. Polska, rządowy projekt ustawy o systemach sztucznej inteligencji (UC71) wraz z konstrukcją Komisji Rozwoju i Bezpieczeństwa Sztucznej Inteligencji — stan legislacyjny wymaga aktualizacji przed publikacją.
132. Wyrok Trybunału Konstytucyjnego w sprawie profilowania osób bezrobotnych (K 53/16) — ranga aktu jako podstawa niezgodności.

E.14. Administracja AI: dowody empiryczne, rejestry, nadzór

133. Badania nad pojemnością nadzoru ludzkiego i odwróconą zależnością bezpieczeństwa od wskaźnika eskalacji, 2025–2026.
134. Badania prerejestrowane nad wpływem nadzoru ludzkiego na wskaźnik błędu w decyzjach o konsekwencjach finansowych, 2025–2026.
135. Quasi-eksperymentalne badania nad odstępowaniem od rekomendacji algorytmu w decyzjach sądowych.
136. Fundacja Panoptykon, analizy profilowania osób bezrobotnych w Polsce.
137. Holenderska Izba Obrachunkowa, badania rejestru algorytmów administracji centralnej.
138. Badania jakościowe nad użytecznością rejestrów algorytmicznych dla organów nadzoru i organizacji strażniczych.
139. Kanadyjski rejestr federalny systemów zautomatyzowanego podejmowania decyzji — analizy pokrycia i „milczeń biurokratycznych”.
140. NASK, badania postaw kadry administracji publicznej wobec sztucznej inteligencji.
141. Najwyższa Izba Kontroli, kontrole bezpieczeństwa informacji w jednostkach samorządu terytorialnego.
142. Sieć Obywatelska Watchdog Polska, badanie jawności użycia systemów AI w instytucjach publicznych.
143. Analizy incydentów AI po wprowadzeniu do obrotu wobec obowiązków monitorowania, 2020–2026.
144. Systematyczny przegląd literatury o wycofywaniu i dekomisji systemów AI w sektorze publicznym.
145. Stanford CRFM, Foundation Model Transparency Index, 2023–2024.
146. Lingjiao Chen, Matei Zaharia, James Zou, How Is ChatGPT’s Behavior Changing Over Time?, 2023 — dryf modelu.
147. Matthew Dahl i in., Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models, 2024 — pozycja przywoływana w destylacie jako cytaty wtórny.
148. Badania nad zamówieniami publicznymi jako narzędziem kształtowania publicznej AI: progi kwotowe, darowizny technologii, umowy przyłączeniowe.

E.15. Suwerenność technologiczna i modele językowe

- 149. PLLuM — polski model językowy rozwijany ze środków publicznych przez konsorcjum instytucji naukowych, 2024–2026.
- 150. Bielik — otwarty polski model językowy rozwijany przez społeczność SpeakLeash, 2024–2026.
- 151. Apertus — w pełni otwarty model wielojęzyczny trenowany na publicznej infrastrukturze obliczeniowej, 2025.
- 152. Open Future, wymagania gwarantowanego wyjścia i multi-sourcingu w zamówieniach publicznych na AI, 2026.
- 153. CERRE, analizy preferencji europejskiej i kosztów suwerenności cyfrowej, 2025–2026.
- 154. Opracowania o roli „szybkiego naśladowcy” (fast follower) w polityce modeli fundamentowych sektora publicznego.
- 155. Materiały o suwerennej infrastrukturze obliczeniowej, kontroli kluczy kryptograficznych i ekspozycji na obce reżimy prawne.

E.16. Wojna informacyjna, manipulacja i odporność poznawcza

- 156. Europejska Służba Działań Zewnętrznych, raporty o obcej manipulacji informacyjnej i ingerencji (FIMI), 2023–2026.
- 157. NATO, opracowania o wojnie poznawczej (cognitive warfare), 2021–2026.
- 158. Sander van der Linden, Jon Roozenbeek, prace o uodpornianiu wyprzedzającym (inoculation, prebunking).
- 159. Badania nad skoordynowanym zachowaniem nieautentycznym i wykrywaniem sieci kont pozorowanych.
- 160. Prace o perswazyjności modeli językowych i o ryzyku syntetycznego konsensusu wytwarzanego przez roje agentów, 2024–2026.
- 161. Richard Thaler, Cass Sunstein, Nudge, Yale University Press 2008 — architektura wyboru i jej nadużycia.
- 162. Shoshana Zuboff, The Age of Surveillance Capitalism, PublicAffairs 2019 — ekonomia uwagi i asymetria danych.

E.17. Demokracja, kontrola obywatelska i technologie obywatelskie

- 163. Materiały o platformach deliberacyjnych i budżecie partycypacyjnym (Polis, vTaiwan, Decidim).
- 164. Prace o AI wspierającej deliberację oraz o granicach zastępowania deliberacji przez systemy generatywne.
- 165. Opracowania o prawie do decyzji człowieka, prawie do wyjaśnienia i prawie do odwołania w postępowaniu administracyjnym.
- 166. Prace o rozliczalności jako obowiązku zdania sprawy (answerability) i o ograniczeniach przejrzystości jako narzędzia zaufania.
- 167. Analizy współprodukcji usług publicznych i roli organizacji obywatelskich jako agentów zmiany.

E.18. Symulacja społeczna i modelowanie agentowe polityk

168. Robert Axelrod, *The Complexity of Cooperation*, Princeton University Press 1997.
169. Joshua M. Epstein, Robert Axtell, *Growing Artificial Societies: Social Science from the Bottom Up*, MIT Press 1996.
170. „Journal of Artificial Societies and Social Simulation” (JASSS) — czasopismo referencyjne modelowania agentowego.
171. Prace o modelach generatywnych społeczeństwa i cyfrowych bliźniakach wspólnot, 2023–2026.
172. Krytyczne analizy zastosowania modeli językowych w modelowaniu agentowym oraz audyty odporności wyników symulacji.

E.19. Destylaty i korpusy operacyjne TechnoRepubliki

173. Metodyka agentów w tworzeniu TechnoRepubliki Polskiej — destylat operacyjny korpusu biblioteki agentowej (15 klastrów).
174. Struktura agentowa TechnoRepubliki Polskiej — destylat strukturalny: warstwy, węzły, krawędzie, pętle sterowania.
175. Rzeczpospolita — system multiagentowy: destylat strategiczno-inżynieryjny korpusu wieloagentowego (osiem składników formuły).
176. Administracja AI TechnoRepubliki — destylat korpusu administracyjnego: dowody empiryczne, prawo, rejestry, zamówienia, dekomisja.
177. Cybernetics of State — destylat o cybernetyce państwa, narodu i organizacji politycznej.
178. Zarządzanie wiedzą poprzez SEO — destylat o dyfuzji wiedzy programu TechnoSanacji.
179. Nota: destylaty stanowią warstwę pośrednią między korpusem źródłowym a dokumentami programowymi; przy publikacji zewnętrznej pozycje instytucjonalne z lat 2023–2026 wymagają uzupełnienia o numery dokumentów i adresy elektroniczne.

Polska, 08.09.2026. - Materiał udostępniony na licencji Creative Commons - Uznanie autorstwa

