

AI Governance dla przedsiębiorcy #Regulacje

Polityka AI · Człowiek w centrum · Zarządzanie ryzykiem AI

dr Michał Nowakowski, r.pr. · legalarchitect.pl / cyberszlif.pl · GovernedAI

Polityka, człowiek i ryzyko - domykamy cykl_

01

Polityka AI

Od tego zaczynamy i na tym kończymy.
Dlaczego polityka zarządzania AI to nie
regulamin korzystania z ChatGPT - i co musi
rozstrzygać.

02

Człowiek w centrum AI

Role i prawa decyzyjne, nadzór, który
naprawdę działa, kompetencje wg art. 4 AI Act
- i co z agentami.

03

Zarządzanie ryzykiem AI

Tylko dla wybranych czy dla każdej
organizacji? Ocena ryzyka i ocena wpływu,
bramki, incydenty, powrót do polityki.

Trzy rezultaty: szkic polityki zarządzania AI, mapa ról i decyzji oraz karta ryzyka jednego zastosowania.

Całość oparta na bibliotece dokumentów AI Governance legalarchitect.pl (ISO/IEC 42001 + AI Act) - skrót na końcu. Wsparcie modelu GPT 6 (Astra) oraz Fable 5.1.

Minimum Viable AI Governance - sześć elementów

1 Rejestr

Co mamy, kto za to odpowiada, na jakich danych.

2 Zasady i narzędzia

DZIŚ

Polityka zarządzania AI + lista dozwolonych narzędzi.

3 Klasyfikacja i ryzyko

DZIŚ

Trzy pytania, poziom ryzyka, ocena i decyzja.

4 Właściciel i bramki

DZIŚ

Jedna osoba na system; decyzja przed startem, zmianą, wycofaniem.

5 Nadzór i incydenty

DZIŚ

Kanał zgłoszeń, prawo zatrzymania, poziomy incydentów.

6 Ludzie

DZIŚ

Kompetencje AI dopasowane do ról.

Dziś pogłębiamy pięć z sześciu elementów. Rejestr jest warunkiem: polityka musi mieć do czego się odnosić.

Firma „Polskie Jabłka”

120 osób. AI w rekrutacji, obsłudze klienta i pracy biurowej. Prezes chce uporządkować zasady i odpowiedzialność - i dowiedzieć się, co właściwie zaakceptował.

Jan Kowalski, prezes zarządu, chce ustalić:

- jakie zastosowania AI firma dopuszcza i na jakich zasadach,
- kto odpowiada za dane, wyniki i nadzór,
- jak oceniać ryzyko i co robić, gdy coś pójdzie nie tak.

„Coś tam zaakceptowałem ludziom z IT i bezpieczeństwa, ale nie wiem, co dokładnie.”

Ten sam model przykładu działa dla usług, handlu, produkcji i organizacji społecznych.

CO JUŻ DZIAŁA W FIRMIE?

- Funkcje AI w CRM/ERP, w tym agent reklamacji
- Chatbot z dokumentami firmy (RAG)
- Moduł AI w systemie rekrutacyjnym (ATS)
- Osobisty agent AI prezesa
- Copilot i ChatGPT na kontach prywatnych



01

MODUŁ 3 /

Polityka AI_

Od tego zaczynamy i na tym kończymy.

Polityka AI to nie regulamin korzystania z ChatGPT_

REGULAMIN KORZYSTANIA Z NARZĘDZI AI

Standard dopuszczalnego użycia

- Kto może używać jakich narzędzi i na jakich kontach
- Jakie dane wolno wkleić (klasy danych)
- Weryfikacja wyników, ujawnianie użycia AI
- Jak zgłosić nowe narzędzie i incydent

Potrzebny. Ale obejmuje tylko ludzi przy klawiaturze. Nie mówi nic o agencie reklamacji w CRM ani o module w ATS.

POLITYKA ZARZĄDZANIA AI (ISO/IEC 42001 KL. 5.2)

Dokument nadrzędny: zasady i prawa decyzyjne

- Obejmuje każdy system AI: kupiony, zbudowany, wbudowany, agenta
- Zasady, role, klasyfikacja, cykl życia, ryzyko, dane, dostawcy
- Kto decyduje, kto akceptuje ryzyko, kto może zatrzymać
- Incydenty, przejrzystość, kompetencje, wyjątki, przegląd

Zatwierdza zarząd. Regulamin narzędzi jest jednym z jej dokumentów wykonawczych.

Test: czy Twoja „polityka AI” mówi coś o systemie, którego nikt z pracowników nie obsługuje ręcznie?

Co musi rozstrzygać polityka zarządzania AI_

01

Zakres i zasady

Kogo obejmuje; 8 zasad, które wiążą każdą decyzję o AI.

02

Role i prawa decyzyjne

Zarząd, komitet AI, AI Governance Officer, Właściciel Systemu, Właściciel Danych.

03

Rejestr i klasyfikacja

Każdy system w rejestrze; trzy pytania; poziom Niski / Średni / Wysoki.

04

Ryzyko i wpływ

Ocena ryzyka dla firmy + ocena wpływu na ludzi; apetyt na ryzyko.

05

Cykl życia i bramki

G0-G5: nic nie idzie na produkcję bez decyzji właściwej jednostki.

06

Dane i dostawcy

Legalność, jakość, retencja; due diligence i klauzule.

07

Dopuszczalne użycie

Narzędzia, konta, klasy danych - czyli ten regulamin, ale jako wykonanie zasady.

08

Incydenty i przejrzystość

Drobny / Istotny / Poważny, 24 h i 72 h; informacja dla użytkowników i osób.

09

Kompetencje

Art. 4 AI Act: szkolenie dopasowane do roli, nadzór przed startem.

10

Przegląd i wyjątki

Wyjątek ma właściciela, termin i środki kompensujące; przegląd roczny.

Polityka ustala reguły i kto decyduje. Jak to zrobić - w standardach, procedurach i instrukcjach.

Jedna polityka, wiele dokumentów wykonawczych_

POZIOM 0

Polityka zarządzania AI · Księga AIMS

Zasady i prawa decyzyjne. Zatwierdza zarząd. Zmiana raz w roku.

POZIOM 1

Standardy i procedury

Cykl życia (sSDLC), dane, dopuszczalne użycie, kontrolki techniczne · ryzyko, ocena wpływu, dostawcy, incydenty, audyt.

POZIOM 2

Instrukcje i szablony

Etyczna AI, agenci, monitorowanie (ML/AIOps) · karta modelu i systemu, szablon AISIA, karta inicjalna projektu.

POZIOM 3

Rejestry, deklaracja, komunikaty

Rejestr systemów, macierz ryzyk, Deklaracja stosowania (SoA), Oświadczenie o przejrzystości.

Start dla małej firmy: polityka (3-5 stron), rejestr, standard użycia narzędzi, procedura ryzyka. Reszta - gdy pojawi się potrzeba.

Osiem zasad - i co z nich wynika w praktyce_

Zasada	Co oznacza w decyzjach
Zgodność z prawem	Praktyki zakazane (art. 5) wykluczone na starcie; wysokie ryzyko = dokumentacja i rejestracja.
Nadzór człowieka	Decyzja o osobie zawsze z realnym przeglądem człowieka; każdy system da się zatrzymać.
Bezpieczeństwo i odporność	Model zagrożeń AI, testy, plan awaryjny, wycofanie zmian - przed startem.
Ochrona danych	Zbiór bez statusu „zatwierdzony” nie trafia do trenowania ani do produkcji.
Sprawiedliwość	Grupy, które system może traktować odmiennie, znane przed G1; progi różnic w kryteriach odbioru.
Przejrzystość	Użytkownik wie, że to AI; osoba dotknięta dostaje wyjaśnienie i drogę odwołania.
Rozliczalność	Jeden właściciel na system; decyzje bramkowe z datą, organem i warunkami.
Proporcjonalność	Wymagania rosną z poziomem ryzyka; minimum obowiązuje zawsze.

Trzy pytania, trzy poziomy, różne wymagania_

Q1 Czy AI podejmuje lub wspiera decyzje o ludziach?

Q2 Czy ludzie wchodzi w interakcję z AI lub dostają jej treści?

Q3 Czy dane trafiają do zewnętrznego dostawcy modelu?

NISKI

Rejestr, właściciel, zasady użycia, bramki G0 i G03 (modelowe), testy funkcjonalne, podstawowe monitorowanie.

ŚREDNI

+ ocena ryzyka, uproszczona ocena wpływu, pełne bramki, model zagrożeń, testy bezpieczeństwa i sprawiedliwości, plan monitorowania, karta systemu.

WYSOKI

+ pełna ocena wpływu, DPIA, niezależne testy, red-teaming co rok, decyzja komitetu AI, dokumentacja wg AI Act, jeżeli wysokie ryzyko prawnie.

Q1 = tak → Wysoki (sugerowany)

Q2 lub Q3 = tak → Średni

wszystko nie → Niski

Poziom to model zarządczy. Kategorię prawną AI Act (zakazane / wysokie ryzyko / art. 50) ustalamy osobno - obie wpisujemy do rejestru.

Jak brzmi klauzula, która działa_

Rejestr (pkt 6.1)

„System AI nieobecny w rejestrze nie jest używany produkcyjnie. Wpis powstaje przed bramką G0 i zawiera właściciela, cel, dane i wstępny poziom ryzyka.”

Da się sprawdzić: lista systemów vs rejestr.

Nadzór (pkt 11.2)

„Wynik AI wpływający na decyzję o skutku prawnym lub podobnie istotnym dla osoby jest przed jej podjęciem jest przeglądany przez człowieka mającego kompetencje i uprawnienie do zmiany decyzji.”

Da się sprawdzić: kto, kiedy, czy mógł zmienić.

Wyjątki (rozdz. 17)

„Wyjątku udziela Komitet AI na czas określony, nie dłuższy niż 12 miesięcy, z udokumentowanymi środkami kompensującymi. Wyjątek nie uchyla wymagań prawa.”

Da się sprawdzić: termin, właściciel, środki.

Każda klauzula ma decydenta albo warunek do sprawdzenia. Bez tego zostaje deklaracja.

Wyjątek ma właściciela, termin i środek kompensujący_

01

Wniosek

Cel, narzędzie, dane, powód odstąpienia, czas użycia. Składa właściciel zastosowania.

02

Ocena

Skutki, obowiązki prawne, dodatkowe zabezpieczenia. AIGO + bezpieczeństwo + ochrona danych, gdy dotyczy.

03

Decyzja

Zgoda z warunkami, odmowa albo ograniczony pilotaż. Komitet AI; do 12 miesięcy.

04

Zamknięcie

Przegląd wyniku, wygaśnięcie zgody, usunięcie dostępów. Wpis w rejestrze wyjątków.

Wyjątek od polityki wewnętrznej nie zwalnia z obowiązków prawnych - i nigdy nie dotyczy praktyk zakazanych.



02

MODUŁ 3 /

Człowiek w centrum AI_

Role, nadzór, kompetencje - i agenci, którzy działają sami.

Kto za co odpowiada - mapa ról

Rola	Decyduje o	W firmie 120 osób
Zarząd	Polityka, apetyt na ryzyko, zasoby, ryzyka krytyczne i wyjątki od zakazów	Prezes + wiceprezes
Komitec AI (AIGC)	Zatwierdza systemy Średnie i Wysokie, ocenę wpływu, dostawców krytycznych, wyjątki	Zarząd + szef IT + osoba od prawa/danych + szef biznesu; spotkanie raz na kwartał
AI Governance Officer (AIGO)	Rejestr, bramki, klasyfikacja, incydenty, raport do zarządu; może zawiesić system	Jedna osoba z pełnomocnictwem (np. szef operacji, compliance, IT) - nie musi być prawnikiem
Właściciel Systemu AI	Cel, wyniki, przeglądy, zmiany, wdrożenie środków - za swój system	Kierownik procesu: HR za ATS, szef obsługi za agenta reklamacji
Właściciel Danych	Legalność, jakość i retencja zbiorów	Osoba odpowiedzialna za CRM / bazę kandydatów
Bezpieczeństwo (ISO) i ochrona danych (IOD)	Model zagrożeń, DPIA, zgoda na dane szczególne	Zewnętrzny IOD + administrator IT
Użytkownik i nadzorujący	Sprawdza wynik, dokumentuje decyzję, zgłasza problem, może zatrzymać	Każdy pracownik; nadzorujący wskazani per system

Role można łączyć. Nie można zostawić odpowiedzialności bez właściciela.

Nadzór człowieka, który naprawdę działa_

- | | | |
|----|----------------------------|---|
| 01 | Rozumie | Zna cel, ograniczenia i tryby błędu systemu; widzi źródła i poziom pewności, nie tylko wynik. |
| 02 | Ma warunki | Czas, dane i wsparcie do sprawdzenia; interfejs bez domyślnego „zatwierdź”. |
| 03 | Może zakwestionować | Odrzuca rekomendację, podejmuje własną decyzję i ją dokumentuje. |
| 04 | Może zatrzymać | Ma uprawnienie i przetestowaną ścieżkę zatrzymania (kill switch w 5 minut dla agentów). |
| 05 | Jest mierzony | Odsetek zmian i odrzuceń, czas decyzji, odsetek akceptacji bez otwarcia - w planie monitorowania. |

Systemy wysokiego ryzyka: art. 14 i 26 ust. 2 AI Act. Pozostałe: ta sama lista jako dobra praktyka.

Sygnal ostrzegawczy z monitorowania: > 50% wyników zaakceptowanych bez otwarcia lub mediana decyzji < 3 sekundy.

Kompetencje AI dopasowane do ról (art. 4 AI Act)_

Kto	Co musi umieć	Forma i częstotliwość
Cały personel	Czym jest AI, ograniczenia (halucynacje, stronniczość), zasady użycia, klasy danych, weryfikacja wyniku, zgłaszanie	E-learning 60-90 min z testem (próg 80%) - przed dostępem do narzędzi, co 12 mies.
Personel nadzoru	Działanie konkretnego systemu, tryby błędu, automatyczna stronniczość, zatrzymanie, obowiązki wobec osób	Szkolenie per system 2-4 h z ćwiczeniem (test wykrywania błędnych wyników) - przed startem i co 12 mies.
Właściciele systemów	Cykl życia i bramki, ocena ryzyka i wpływu, dostawcy, karty, monitorowanie, incydenty	Kurs 2 dni + egzamin - przed objęciem roli, co 24 mies.
Zespół techniczny	Model zagrożeń AI, wstrzyknienia, izolacja agentów, ewaluacja, logowanie	Warsztat 2 dni + ćwiczenia red-team - co 12 mies.
AIGO, komitet, audytorzy	ISO/IEC 42001, 42005, EN 18286, AI Act szczegółowo, RODO w AI, techniki audytu	Szkolenie zewnętrzne lub certyfikacja; aktualizacje co kwartał
Zarząd	Obowiązki i odpowiedzialność, profil ryzyka, decyzje wymagane od zarządu	Briefing 2 h rocznie + przed decyzją o wyjątku

Dowód kompetencji nadzoru jest warunkiem bramki G3: bez przeszkolonych ludzi system nie rusza.

Przykład HR: wpis w rejestrze systemów AI_

01 NAZWA I CEL

moduł AI w ATS ocenia CV i sugeruje, kogo zaprosić

03 ŹRÓDŁO

funkcja dostawcy ATS, konfigurowana przez HR → firma = podmiot stosujący

05 UŻYTKOWNICY / OSOBY DOTKNIĘTE

HR / kandydaci (ok. 1 200 rocznie)

07 TRZY PYTANIA / POZIOM

Q1 tak, Q2 nie, Q3 tak → Wysoki · AI Act: Zał. III pkt 4 lit. a - wysokie ryzyko

02 WŁAŚCICIEL

kierownik HR (Właściciel Systemu AI)

04 DOSTAWCA I DANE

dostawca ATS z UE, model zewnętrzny · CV, historia zatrudnienia, wyniki rozmów

06 ETAP

8 miesięcy w użyciu, bez oceny przed startem (brak G0-G3)

08 OSTATNI PRZEGLĄD / LUKI

nigdy · brak oceny ryzyka, AISIA, DPIA, szkolenia nadzoru

Sam wpis w rejestrze już pokazuje, czego brakuje. To pierwszy krok, nie ostatni.

HR: trzy pytania i co z nich wynika_

1. Czy AI podejmuje lub wspiera decyzje o ludziach?

TAK

2. Czy kandydaci rozmawiają z AI lub dostają treści z AI?

NIE

3. Czy dane kandydatów trafiają do zewnętrznego modelu?

TAK

WYNIK

AI Act: ranking kandydatów = wysokie ryzyko (Zał. III pkt 4). Firma jest podmiotem stosującym (art. 26): nadzór, dane wejściowe, logi, informacja dla kandydatów, instrukcja dostawcy. Termin stosowania: 2.12.2027 (po Digital Omnibus).

RODO: DPIA (art. 35), zakaz decyzji wyłącznie automatycznej (art. 22) - udział HR musi być rzeczywisty.

Governance: poziom Wysoki → pełna ocena ryzyka i wpływu (AISIA), bramki, przeszkolony nadzór, decyzja komitetu AI.

**Rzeczywisty nadzór HR
przed decyzją**

Minimum: HR rozumie wynik, sprawdza jego podstawy i może go odrzucić. Sama akceptacja rankingu nie wystarcza.

Stan prawny i terminy sprawdzone 17.10.2026. Przepisy przejściowe dla już używanego systemu sprawdzamy oddzielnie (art. 111).

Decyzja rekrutera_

AI odrzuca doświadczoną kandydatkę. Uzasadnienie w systemie: „niski potencjał adaptacji”.

Co sprawdzasz przed decyzją?

- A** Jakie dane i kryteria wpłynęły na wynik? Czy wśród nich jest wiek, płeć, przerwa w zatrudnieniu?
- B** Czy możesz odrzucić wynik i uzasadnić własną ocenę - i czy ktoś to zapisze?
- C** Komu zgłosisz podejrzenie dyskryminacji - i co stanie się z systemem po zgłoszeniu?

W bibliotece: instrukcja etycznej AI, rozdz. 7 - decyzje dotyczące osób; procedura incydentów - kanał zgłoszeń.

Agent działa sam - więc człowiek wyznacza granice_

A0

Asystujący

Agent proponuje; każde działanie wykonuje człowiek.

Zatwierdza: Właściciel systemu

A1

Z potwierdzeniem

Agent działa po każdorazowym zatwierdzeniu; odczyt automatyczny.

Zatwierdza: Właściciel + AIGO

A2

Ograniczony

Działa sam z zamkniętej listy w limitach; działania istotne wymagają zgody.

Zatwierdza: Komitet AI

A3

Autonomiczny

Szeroki zakres, nadzór po fakcie. Tylko bez decyzji o osobach i skutków nieodwracalnych.

Zatwierdza: Komitet AI + zarząd

MANIFEST AGENTA (POLICY-AS-CODE)

Cel i zakres · poziom autonomii · dozwolone narzędzia z parametrami (np. e-mail: tylko domena wewnętrzna, max 5 adresatów) · działania istotne wymagające zatwierdzenia (płatności, komunikacja na zewnątrz, zmiany w CRM/HR, wszystko nieodwracalne) · limity kroków, czasu i kosztu · pamięć i retencja · kill switch (test co 6 mies.) · właściciel.
Działanie spoza manifestu = blokada + incydent.

Agent podejmujący lub przygotowujący decyzje o osobach: maksymalnie A1.

Governance agenta: od mandatu do operacji

Przykład: agent obsługi reklamacji z dostępem do danych i CRM.

MANDAT ZATWIERDZONY PRZEZ WŁAŚCICIELA

Cel · tożsamość · dozwolone dane i operacje · limity · osoby zatwierdzające · termin przeglądu

Zgłoszenie / RAG

Treść zewnętrzna. Nie nadaje uprawnień.



Agent i model

Interpretuje zadanie. Proponuje operację.



Kontrola wykonania

Brama narzędzi: dostęp, reguły, wymagana zgoda.



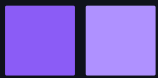
CRM / narzędzie

Wykonuje wyłącznie dopuszczoną operację.

NADZÓR I ZATRZYMANIE

Logi całej ścieżki (kto, co, kiedy, na jakiej podstawie) · alerty · kill switch: odebranie tokenów i blokada dalszych operacji w ≤ 5 minut

Decyzja modelu jest propozycją działania. Uprawnienia egzekwuje niezależna warstwa - nie prompt.



03

MODUŁ 3 /

Zarządzanie ryzykiem AI_

Tylko dla wybranych czy dla każdej organizacji?

Dla każdej organizacji - w skali dopasowanej do skutków_

NISKI

Proste wsparcie

Publiczne treści i odwracalny wynik. Krótka ocena przesiewowa (lista kontrolna), kontrola jakości, jasne zasady użycia.

ŚREDNI

Dane i integracje

Informacje klientów lub dostęp do systemów. Ocena ryzyka, uproszczona ocena wpływu, ocena dostawcy, scenariusze nadużyć, plan monitorowania.

WYSOKI

Istotne decyzje

Wpływ na ludzi, pieniądze lub bezpieczeństwo. Pełna ocena ryzyka i wpływu, testy niezależne, nadzór, akceptacja ryzyka przez komitet AI.

Obowiązek prawny vs dobra praktyka: system zarządzania ryzykiem z art. 9 AI Act obowiązuje dostawcę systemu wysokiego ryzyka. Ocena ryzyka każdego istotnego zastosowania to wymaganie ISO/IEC 42001 i zdrowy rozsądek - brak kwalifikacji „wysokie ryzyko” nie oznacza braku ryzyka.

Dwie oceny, jedna macierz ryzyk

OCENA RYZYKA AI

Ryzyko dla organizacji

- Zgodność i prawo, jakość, bezpieczeństwo, dane, dostawcy, ciągłość, reputacja
- Przed G1 (wstępna), przed G3 (pełna), przy G4 i wyzwalaczach
- Właściciel: Właściciel Systemu AI; zatwierdza AIGO / komitet AI
- ISO/IEC 42001; ISO/IEC 23894; AI Act art. 9

OCENA WPŁYWU (AISIA)

Wpływ na ludzi i społeczeństwo

- Prawa podstawowe, zdrowie, finanse osób, równość, prywatność, autonomia, dostęp do usług, skutki społeczne
- Przesiew przy G0 dla każdego; pełna dla Średnich i Wysokich
- Zawiera elementy DPIA (RODO art. 35) i FRIA (AI Act art. 27)
- ISO/IEC 42005; ISO/IEC 42001 kl. 6.1.4

JEDNA MACIERZ RYZYK · TE SAME SKALE

Prawdopodobieństwo 1-5 × skutek 1-5. Poziom: Niski 1-4 · Średni 5-9 · Wysoki 10-14 · Krytyczny 15-25. Każdy wpływ z AISIA oceniony jako Średni lub wyższy jest ryzykiem w macierzy z właścicielem i działaniem.

Mała firma: jedna tabela w Excelu wystarczy na start. Liczy się scenariusz, właściciel i dowód działania kontroli.

Ocena ryzyka w siedmiu krokach - i apetyt zarządu

			SKUTEK ↓ / PRAWDOPODOBIEŃSTWO →					
01	Kontekst	Cel, dane, użytkownicy, osoby dotknięte, dostawca, autonomia.	5	5	10	15	20	25
02	Scenariusz	Przyczyna - zdarzenie - skutek. Katalog źródeł ryzyka AI.	4	4	8	12	16	20
03	Analiza	Prawdopodobieństwo × skutek przed środkami. W razie sporu wyżej.	3	3	6	9	12	15
			2	2	4	6	8	10
04	Ewaluacja	Porównanie z apetytem: co wymaga postępowania.	1	1	2	3	4	5
05	Postępowanie	Kontrolka, właściciel, termin, bramka. Ryzyko rezydualne.		1	2	3	4	5
06	Decyzja	AIGO / komitet AI / zarząd - zależnie od poziomu.						
07	Przegląd	KRI w monitorowaniu; po incydencie, zmianie, co 12 mies.						

Reguły dodatkowe: praktyka zakazana = Krytyczne z definicji. Skutek 5 = co najmniej Wysokie. Osoby podatne podnoszą skutek o 1. Niezgodności z prawem nie akceptujemy, bo „wyszło mało punktów”.

HR: od scenariusza szkody do kontroli_

WARTOŚĆ I HACZYK

Wartość: uporządkowanie 1 200 CV rocznie i mniej pracy przy selekcji.

Haczyk: dyskryminacja ze względu na wiek, płeć lub narodowość (dane historyczne odzwierciedlają wcześniejsze decyzje). Automatyczne odrzucenie bez oceny HR. Brak dokumentacji utrudnia obronę w sporze.

RYZYO NR 1 · PRZED ŚRODKAMI

15

3×5 · Krytyczne

Ranking niesprawiedliwie traktuje grupy kandydatów. Prawdopodobieństwo 3 (dane historyczne, brak testów) $\times$ skutek 5 (systematyczna dyskryminacja, kara, szkoda dla osób).

Drugie ryzyko: brak podstawy prawnej i informacji dla kandydatów: $3 \times 4 = 12$, Wysokie.

RODO i prawo pracy

KONTROLKI I RYZYKO REZYDUALNE

Ważenie danych i test sprawiedliwości wg grup (próg 3 p.p.) · ranking tylko jako podpowiedź, decyzję podejmuje rekruter · informacja dla kandydatów i droga odwołania · DPIA + FRIA · szkolenie nadzoru HR.

Po środkach: $2 \times 4 = 8 \rightarrow$ Średnie. Akceptuje komitet AI przed G3.

Skala i próg to założenia przykładowe. Brak przekroczenia progu nie dowodzi braku dyskryminacji - decyduje właściciel na podstawie dowodów.

Wpis w macierzy ryzyk: agent obsługi reklamacji_

Pole	Przykładowy wpis
Scenariusz (przyczyna - zdarzenie - skutek)	Instrukcja ukryta w treści zgłoszenia - agent uznaje niezasadną reklamację i wysyła potwierdzenie - strata finansowa, spór z klientem, utrata zaufania.
Właściciel / kategoria / źródło	Kierownik obsługi klienta · Bezpieczeństwo · ocena ryzyka przed G2
Ocena inherentna	Prawdopodobieństwo 4 × skutek 3 = 12 → Wysokie
Postępowanie (modyfikacja)	Manifest A2: limit 2 000 zł, potwierdzenie wysyłki i uznania, treść zgłoszenia = dane, brama narzędzi, pełne logi, kill switch. Właściciel działania: lider techniczny; termin: przed G3.
Dowód skuteczności	Test przekroczenia limitu, próby obejścia zatwierdzenia i wstrzyknięcia (0 działań spoza manifestu); data, wersja, osoba.
Ryzyko rezidualne / akceptacja	$2 \times 3 = 6 \rightarrow$ Średnie · komitet AI, data · KRI: działania spoza manifestu > 0 = incydent
Przegląd	Po zmianie narzędzia lub modelu, po incydencie, co 12 miesięcy (A2: co 6)

Decyzję opieramy na dowodach z testu, nie na samym wpisie w rejestrze.

Kontrola wymaga testu - i dowodu_

01

Dane i jakość

Zbiór wzorcowy (golden dataset): błędy na reprezentatywnych przypadkach, także nietypowych; różnice między grupami; halucynacje w RAG.

02

Uprawnienia

Spróbuj przekroczyć limit, wywołać niedozwoloną operację, ominąć zatwierdzenie, wstrzyknąć polecenie przez dokument. Wynik oczekiwany: 0.

03

Ludzie i reakcja

Przećwicz rozpoznanie problemu, zatrzymanie w 5 minut, pracę w trybie awaryjnym i zgłoszenie incydentu. Ćwiczenie stołowe raz w roku.

Dowód: wynik testu, data, wersja systemu, osoba odpowiedzialna, usunięte błędy. Bez tego kontrolka „istnieje” tylko w dokumencie.

Decyzja o uruchomieniu i zmianie: bramki G0-G5



Decyzja	Kiedy
Zgoda	Kontrole działają, ryzyko ma właściciela, nadzór przeszkolony, luki P1 = 0.
Zgoda warunkowa	Ograniczony pilotaż, dodatkowa kontrola i termin ponownej decyzji (np. raport sprawiedliwości po 90 dniach).
Zwrot / odrzucenie	Brak podstawy prawnej, praktyka zakazana, ryzyko Krytyczne, nieskuteczne zabezpieczenie.

Poziom Niski: G0, G3, G5. Zmiana znacząca cofa do G2/G3, zmiana istotna - do G1.

Gdy coś pójdzie nie tak: trzy poziomy, dwa zegary_

Poziom	Kiedy	Terminy wewnętrzne	Decyduje
Drobny	Błędny wynik bez szkody wychwycony przez nadzór; alert ostrzegawczy; próba agenta zablokowana przez bramę	Rejestracja w 5 dni roboczych; zamknięcie w 30 dni	Właściciel systemu
Istotny	Szkoda możliwa do naprawienia; wynik dyskryminujący potwierdzony; udane wstrzyknięcie bez skutku nieodwracalnego; działanie agenta spoza manifestu	Zgłoszenie do AIGO w 24 h; ograniczenie w 72 h; raport w 10 dni roboczych	AIGO
Poważny	Szkoda dla zdrowia lub praw osoby; dane K3 lub osobowe na dużą skalę; decyzje o osobach na błędnych wynikach; nieudane zatrzymanie agenta; praktyka zakazana	AIGO i zarząd natychmiast, max 24 h; zespół w 4 h; raport w 30 dni	AIGO + prawnik; zarząd informowany

DRUGI ZEGAR: TERMINY PRAWNE BIEGNĄ OSOBNO

AI Act art. 73 (dostawca systemu wysokiego ryzyka): poważny incydent do organu nadzoru rynku max 15 dni; śmierć osoby - 10 dni; powszechne naruszenie lub infrastruktura krytyczna - 2 dni. Art. 26 ust. 5 (podmiot stosujący): informacja do dostawcy, zawieszenie użycia. RODO art. 33: naruszenie danych do UODO w 72 h.

Każdy z: właściciel, AIGO, bezpieczeństwo, nadzorujący - może zatrzymać system samodzielnie, gdy zwłoka grozi szkodą.

Wnioski z ryzyka wracają do polityki_

01

Co się wydarzyło?

Błąd, skarga, incydent, nowe zastosowanie, zmiana dostawcy, zmiana prawa (np. Digital Omnibus).

02

Czego zabrakło?

Zasady, roli i uprawnienia, kompetencji czy skutecznej kontrolki technicznej.

03

Co zmieniamy - i na jakim poziomie?

Polityka (zarząd, raz w roku) · standard lub procedura (komitet AI) · instrukcja, lista narzędzi, manifest, konfiguracja (AIGO / właściciel).

04

Po czym poznamy efekt?

Test po zmianie, potwierdzenie wdrożenia, wskaźnik w monitorowaniu, kolejny przegląd.

Od tego zaczynamy i na tym kończymy: polityka żyje, bo zasila ją praktyka.

Komplet dokumentów, o których dziś mówiliśmy_

POLITYKA I KSIĘGA

- **Polityka zarządzania AI ★**
- Księga systemu zarządzania AI (AIMS)

STANDARDY

- **Dopuszczalne użycie narzędzi AI ★**
- Cykl życia systemu AI (sSDLC)
- Zarządzanie danymi AI
- Katalog 58 kontrolek technicznych

PROCEDURY

- **Zarządzanie ryzykiem AI ★**
- Ocena wpływu (AISIA) + szablon
- Dostawcy i klienci
- Incydenty AI
- Kompetencje, audyt, przegląd

INSTRUKCJE I SZABLONY

- Etyczna AI
- Agenci AI (A0-A3, manifest)
- Monitorowanie ML/AIOps
- Karta modelu i systemu
- Karta inicjalna projektu (ROI)

REJESTRY I KOMUNIKACJA

- **Rejestr systemów AI ★**
- Macierz ryzyk
- Deklaracja stosowania (SoA)
- Oświadczenie o przejrzystości

★ start dla małej firmy: polityka, standard użycia narzędzi, procedura ryzyka, rejestr. Wszystko w tych samych rolach, poziomach i bramkach.

Co zrobić w ciągu 30 dni_

01

Właściciel i zakres

Wybierz jedno zastosowanie.
Wpisz je do rejestru: cel, dane,
trzy pytania, poziom, osoba
odpowiedzialna.

02

Polityka, nie regulamin

Sprawdź, czy Twoja „polityka
AI” rozstrzyga role, klasyfikację,
bramki i wyjątki. Jeśli nie -
uzupełnij o decyzje, nie o
deklaracje.

03

Ludzie i nadzór

Wskaż nadzorującego,
przećwicz odrzucenie wyniku i
zatrzymanie. Zaplanuj szkolenie
art. 4 dla wszystkich.

04

Ryzyko i test

Zapisz najważniejszy scenariusz
szkody, kontrolkę, właściciela,
termin. Przetestuj ją i wpisz
wynik. Przenieś wnioski do
polityki.

Jedno zastosowanie, jeden właściciel, jedna karta ryzyka, jeden test. Reszta przyjdzie z praktyki.

Trzy rzeczy do zabrania ze spotkania_

01

Polityka

Polityka zarządzania AI ustala zasady i prawa decyzyjne dla każdego systemu - regulamin narzędzi jest jej wykonaniem, nie zamiennikiem.

02

Człowiek

Jeden właściciel na system, nadzór, który może zrozumieć, odrzucić i zatrzymać, kompetencje dopasowane do roli - także wobec agentów.

03

Ryzyko

Dla każdej organizacji, proporcjonalnie: scenariusz, właściciel, kontrolka, test, decyzja na bramce - i wniosek z powrotem do polityki.

Pierwszy krok: jedna karta zastosowania i rozmowa z jego właścicielem.

Od czego zacznie Twoja organizacja?_

A Od polityki zarządzania AI - i sprawdzenia, czy to nie jest tylko regulamin narzędzi.

B Od właścicieli, Komitetu AI i kompetencji ludzi.

C Od oceny ryzyka jednego zastosowania - i jednego testu.

Odpowiedź na czacie: litera + jedna osoba + data.

Dziękuję_

Polityka, człowiek, ryzyko - i jedna karta zastosowania na początek.



AI I PRAWO

legalarchitect.pl

CYBERBEZPIECZEŃSTWO

cyberszlif.pl

GOVERNANCE JAKO USŁUGA

GovernedAI

KONTAKT

michal@nowakowski.ai ·
linkedin.com/in/mjnowakowski