Liderzy.AI

Od procesów **AI-Ready** do **AI-First**

Modele projektowania procesów biznesowych

w erze sztucznej inteligencji

prof. SGH, dr hab. Andrzej Sobczak
Kierownik Zakładu Zarządzania IT
Szkoła Główna Handlowa w Warszawie

SPIS TREŚCI

SPIS TREŚCI 1
STRESZCZENIE 3
Kluczowe wnioski 3
1. WPROWADZENIE: KONTEKST I GENEZA DWÓCH MODELI PROCESÓW 5
2. KLUCZOWE KONCEPCJE 6
2.1. Proces AI-Ready 6
2.2. Proces AI-First 7
2.3. Rozróżnienie pojęć: czym AI-First nie jest 8
2.4. Pojęcia techniczne: model, agent, system AI, automatyzacja 8
3. ANATOMIA PORÓWNAWCZA PARADYGMATÓW PROCESOWYCH 10
3.1. interpretacja wybranych wymiarów 11
Architektura informacyjna 11
Mechanizm decyzyjny i nadzór człowieka 11
Eskalacja: wielokryteriowa ocena sygnałów zamiast pojedynczego progu 11
Skalowalność i koszt jednostkowy 12
Wyszukiwanie i pobieranie kontekstu (RAG): zakotwiczenie nie jest gwarancją prawdy 12
4. MACIERZ DECYZYJNA: KIEDY STOSOWAĆ DANY PARADYGMAT 13
4.1. Model ekonomiczny zamiast progu wolumenu 14
4.2. Wstępny filtr decyzyjny 14
5. TYPOLOGIA NADZORU CZŁOWIEKA I TYPÓW INTERAKCJI CZŁOWIEK-AI 16
6. ŚCIEŻKI TRANSFORMACJI 19
6.1. Ścieżka adaptacyjna: As-Is → AI-Augmented → AI-Ready 19
Brama 1: As-Is → AI-Augmented 19
Brama 2: AI-Augmented → AI-Ready 20
6.2. Ścieżka rekonstrukcyjna: AI-Ready → decyzja architektoniczna → AI-First 20
Brama 3: warunki uruchomienia ścieżki rekonstrukcyjnej 20
7. RYZYKA I MECHANIZMY ICH OGRANICZANIA 22
7.1. Ryzyka klasyczne wdrożeń AI 22
7.2. Ryzyka bezpieczeństwa agentowego 23
8. PRZYKŁAD PROCESU: OBSŁUGA REKLAMACJI KLIENTA 25
8.1. Komentarz interpretacyjny 26
8.2. Ekonomika trzech wariantów 28
9. KARTA JEDNOSTKI PRACY POZNAWCZEJ 29
10. MINIMALNY STANDARD PROCESU AI-READY 31
11. KIEDY NIE STOSOWAĆ PARADYGMATU AI-FIRST 32
12. RAMY REGULACYJNE I STANDARDY 33

12.1. Praktyczne implikacje dla projektowania procesów 34
12.2. Role regulacyjne w procesach AI-Ready i AI-First 35
13. REKOMENDACJE DLA LIDERÓW TRANSFORMACJI 36
13.1. Rekomendacje strategiczne 36
3.2. Rekomendacje operacyjne 36
13.3. Rekomendacje kompetencyjne 37
14. PODSUMOWANIE 38
SŁOWNIK KLUCZOWYCH POJĘĆ 39
BIBLIOGRAFIA (WYBRANE POZYCJE) 42

STRESZCZENIE

Popularyzacja generatywnej sztucznej inteligencji w zarządzaniu procesami biznesowymi doprowadziła do powstania dwóch alternatywnych podejść ich projektowania: **AI-Ready** oraz **AI-First**. Różnica między nimi ma charakter fundamentalny, ponieważ determinuje odmienne decyzje architektoniczne, modele operacyjne, struktury kosztów, profile ryzyka oraz wzorce ładu zarządczego.

Proces AI-Ready to istniejący proces biznesowy przygotowany do bezpiecznego, mierzalnego i kontrolowanego wykorzystania komponentów AI (modeli lub agentów) przez ludzi-operatorów. Czynności poznawcze w procesie mają opisane dane wejściowe, dane wyjściowe, kryteria jakości, reguły decyzyjne, źródła wiedzy, ścieżki wyjątków, mechanizmy eskalacji oraz wymagania audytowe. Proces AI-Ready nie musi jeszcze wykorzystywać AI. Musi jednak mieć taką strukturę informacyjną i organizacyjną, by wdrożenie AI nie wymagało późniejszego porządkowania wiedzy, danych i modeli odpowiedzialności.

Proces AI-First to proces zaprojektowany przy założeniu, że komponent AI – model, agent lub system wieloagentowy – jest domyślnym kandydatem do wykonania pracy poznawczej. Komponent ten może być osadzony w przepływie decyzyjnym łączącym możliwości AI, reguły deterministyczne oraz integracje systemowe, o ile nie wykluczają tego profil ryzyka, regulacje, koszty błędu, wymagania audytowe lub potrzeba eksperckiego osądu. Człowiek nie znika z procesu, lecz jego rola ewoluuje od wykonywania rutynowych czynności do projektowania zasad działania, nadzoru, recenzji, obsługi wyjątków, kalibracji parametrów jakościowych oraz odpowiedzialności za ryzyko.

Niniejsze opracowanie prezentuje ramy pojęciowe obu paradygmatów oraz ich zestawienie w **ośmiu wymiarach** architektonicznych. Dokument precyzuje także kryteria decyzyjne wyboru właściwego podejścia, czteropoziomową typologię nadzoru człowieka oraz dwie odrębne ścieżki transformacji – adaptacyjną do AI-Ready i rekonstrukcyjną do AI-First. Ponadto skatalogowano w nim kluczowe ryzyka, minimalny standard procesu AI-Ready oraz kryteria wykluczające stosowanie AI-First.

Dokument rozróżnia też dwa poziomy współpracy człowieka z AI: tryby nadzoru człowieka nad decyzją lub akcją systemową oraz operacyjne typy interakcji człowiek-AI stosowane w AI-Ready Process Canvas. Rozróżnienie to pozwala uniknąć mylenia kategorii kontroli i odpowiedzialności z kategoriami projektowania pracy i orkiestracji procesu. Całość osadzono w ramach regulacyjnych obejmujących AI Act, RODO, ISO/IEC 42001, NIST AI Risk Management Framework oraz rozporządzenie DORA dla sektora finansowego.

KLUCZOWE WNIOSKI

- **AI-Ready i AI-First nie są etapami dojrzałości procesowej**. Stanowią dwie odrębne strategie projektowe oraz dwie odrębne ścieżki transformacji: **adaptacyjną** – prowadzącą od procesu klasycznego do AI-Ready, oraz **rekonstrukcyjną** – stosowaną w przypadku procesów istniejących zwykle po osiągnięciu dojrzałości AI-Ready, a w przypadku procesów projektowanych od podstaw –wymagającą wbudowania standardów AI-Ready od początku.
- **Większość organizacji powinna zaczynać od ścieżki adaptacyjnej**. Pozwala to wypracować dyscyplinę procesową, jakość danych i dojrzałość ładu zarządczego bez narażania się na ryzyko nieprzygotowanych wdrożeń AI-First.

- **Ścieżka rekonstrukcyjna jest właściwym wyborem tam, gdzie oczekiwana wartość bieżąca netto (NPV) przejścia do wariantu AI-First jest dodatnia względem scenariusza bazowego** – z uwzględnieniem wolumenu, kosztu jednostkowego, kosztu błędu, kosztu nadzoru i wymogów regulacyjnych. Przesłanką do transformacji nie jest sam wolumen przetwarzanych spraw, lecz dodatnia oczekiwana wartość NPV przedsięwzięcia w analizowanym horyzoncie czasowym.
- **Kluczowym czynnikiem sukcesu jest jakość kontekstu** dostarczanego modelowi. Bez uporządkowanej, wersjonowanej i dostępnej bazy wiedzy oba paradygmaty pozostaną w fazie prototypu.
- **Eskalacja w procesach AI-First nie powinna opierać się wyłącznie na wskaźniku ufności zwracanym przez model**, lecz na wielokryteriowej ocenie sygnałów: klasy ryzyka, kompletności danych, jakości źródeł, zgodności z regułami, wartości sprawy i wykrytych anomaliach.
- **Proces AI-First wymaga dedykowanych ram zarządzania (*AI governance*)** – rejestru modeli, polityki danych, ram oceny ryzyka modelowego, dedykowanej procedury obsługi incydentów oraz adekwatnego nadzoru człowieka, proporcjonalnego do klasyfikacji ryzyka systemu w rozumieniu AI Act.

1. WPROWADZENIE: KONTEKST I GENEZA DWÓCH MODELI PROCESÓW

Pojęcia AI-Ready oraz AI-First stanowią odpowiedzi środowiska analityków, architektów i projektantów procesów na fundamentalne pytanie: w jaki sposób włączyć komponenty AI – w tym modele językowe, klasyfikatory, agentów i systemy wspierające decyzje – do pracy organizacji w sposób bezpieczny, mierzalny i ekonomicznie uzasadniony.

Pierwszą rynkową reakcją na upowszechnianie się generatywnej AI było wdrażanie rozwiązań punktowych – włączanie funkcji AI w istniejące narzędzia bez zmiany procesu (asystenci w pakietach biurowych, klasyfikatory zgłoszeń, generatory streszczeń). Pokazało ono potencjał technologii, lecz jednocześnie ujawniło ograniczenia: wartość była rozproszona, ład zarządczy fragmentaryczny, a wskaźniki efektywności trudne do zagregowania.

W odpowiedzi pojawiła się koncepcja procesu **AI-Ready** – celowej adaptacji procesów, by były one przygotowane do systemowego włączenia komponentów AI. Proces nie powinien być traktowany wyłącznie jako statyczna mapa czynności. Powinien być zarządzany jako zasób operacyjny obejmujący dane, reguły decyzyjne, dokumentację decyzji, odpowiedzialności i ścieżki obsługi wyjątków.

Równolegle w organizacjach o dużym wolumenie powtarzalnej pracy poznawczej oraz w nowych przedsięwzięciach cyfrowych zaczęło kształtować się podejście na wyższym stopniu zaawansowania: projektowanie procesu od podstaw lub jego głęboka rekonstrukcja, przy założeniu, że to model lub agent – a nie człowiek – jest domyślnym wyborem do wykonania czynności poznawczej. W ten sposób ukształtowało się podejście określane w niniejszym opracowaniu jako **AI-First**.

Sformułowano w nim tezę, według której oba paradygmaty mają charakter **komplementarny, a nie konkurencyjny**. Decyzja wdrożeniowa nie sprowadza się do wyboru „uniwersalnie lepszej metody", lecz do dopasowania podejścia do charakterystyki procesu, dojrzałości danych, profilu ryzyka, wymagań regulacyjnych oraz tolerancji organizacji na zmianę.

2. KLUCZOWE KONCEPCJE

2.1. PROCES AI-READY

Definicja

Proces AI-Ready to istniejący proces biznesowy przygotowany do bezpiecz-nego, mierzalnego i kontrolowanego wykorzystania komponentów AI przez ludzi-operatorów. Oznacza to, że czynności poznawcze w procesie mają opisane dane wejściowe, dane wyjściowe, kryteria jakości, reguły decyzyjne, źródła wiedzy, ścieżki wyjątków, mechanizmy eskalacji oraz wymagania audytowe.

Proces AI-Ready nie musi wykorzystywać AI produkcyjnie; musi jednak mieć taką strukturę informacyjną i organizacyjną, by wdrożenie AI nie wymagało późniejszego porządkowania wiedzy, danych i modeli odpowiedzialności.

Proces AI-Ready charakteryzuje się następującymi atrybutami:

- **Zdefiniowane jednostki pracy poznawczej.** Każdy etap procesu wymagający wnioskowania, oceny lub generowania treści ma jasno określone dane wejściowe (dane źródłowe, dokumenty, kontekst), dane wyjściowe (decyzja, dokument, klasyfikacja) oraz kryteria akceptacji.
- **Ustrukturyzowana baza wiedzy.** Wiedza referencyjna procesu – polityki, słowniki, przykłady, precedensy – jest dostępna w formie umożliwiającej automatyczne wyszukiwanie kontekstowe, przy zachowaniu hierarchii wiarygodności oraz mechanizmu wersjonowania źródeł.
- **Zdefiniowane reguły decyzyjne.** Decyzje w procesie są opisane w sposób jawny w formie tabel decyzyjnych, drzew lub polityk, co umożliwia walidację rekomendacji modelu w odniesieniu do stałych kryteriów.
- **Czytelne ścieżki eskalacji.** Dla każdego etapu wspieranego AI istnieje opisana ścieżka przekazania sprawy operatorowi, recenzentowi lub właścicielowi decyzji oraz kryteria jej uruchomienia.
- **Logowanie i audytowalność.** Jeżeli komponent AI jest używany, jego istotne wywołania powinny być rejestrowane zgodnie z polityką audytu, minimalizacji zakresu przetwarzanych danych, retencji i kontroli dostępu – w zakresie niezbędnym do odtworzenia podstawy rekomendacji lub decyzji.

Fundamentalnym rozróżnieniem jest to, że proces AI-Ready **nie wymaga, by AI była w nim aktualnie używana**. Wymaga jedynie, by była gotowa do wdrożenia bez konieczności późniejszego porządkowania danych, dokumentacji i polityk.

2.2. PROCES AI-FIRST

Definicja

Proces AI-First to proces zaprojektowany przy założeniu, że komponent AI – model, agent lub system wieloagentowy – jest domyślnym kandydatem do wykonania pracy poznawczej. Komponent ten może być osadzony w przepływie decyzyjnym łączącym komponenty AI, reguły deterministyczne oraz integracje systemowe, o ile nie wykluczają tego profil ryzyka, regulacje, koszt błędu, wymagania audytowe lub potrzeba eksperckiego osądu.

Człowiek nie znika z procesu, lecz jego rola ewoluuje od wykonywania rutynowych czynności do projektowania zasad działania, nadzoru, recenzji, obsługi wyjątków, kalibracji parametrów jakościowych oraz odpowiedzialności za ryzyko.

Architektura procesu AI-First może zostać zaprojektowana od podstaw albo być wynikiem głębokiej rekonstrukcji procesu istniejącego. Punktem wyjścia jest pytanie o cel procesu (jaki wynik ma zostać dostarczony, na podstawie jakich danych, w jakich ramach jakościowych) – bez założenia sekwencji czynności wynikającej z dotychczasowego sposobu pracy.

Proces AI-First charakteryzuje się następującymi atrybutami:

- **Komponent AI jako domyślny wybór.** Każdy etap pracy poznawczej jest analizowany pod kątem możliwości wykonania przez komponent AI – model, klasyfikator, agenta lub system wieloagentowy – ewentualnie wspierany przez deterministyczne reguły i integracje systemowe. Ostateczny wybór zależy od profilu ryzyka, wymogów regulacyjnych, audytowalności i kosztu błędu.
- **Człowiek w precyzyjnie wyznaczonych punktach kontroli.** Obecność człowieka w procesie wynika z jednej z czterech ról: projektanta (definiuje proces), recenzenta (zatwierdza decyzje o znaczącym wpływie biznesowym), arbitra eskalacji (rozstrzyga sprawy nietypowe) oraz właściciela ryzyka (odpowiada za jakość i zgodność regulacyjną całości).
- **Architektura kontekstu jako fundament.** Centralnym artefaktem przestaje być wyłącznie mapa procesu. Równorzędne znaczenie zyskuje warstwa zasilania modelu danymi: sposób pobierania informacji z bazy wiedzy, mechanizm wywoływania narzędzi zewnętrznych oraz przekazywania ustrukturyzowanych wyników do systemów docelowych.
- **Polityka eskalacji oparta na wielokryteriowej ocenie sygnałów.** Decyzja o tym, czy sprawa kończy się autonomicznie, czy trafia do recenzji, jest podejmowana na podstawie kilku sygnałów: klasy ryzyka sprawy, kompletności danych wejściowych, jakości źródeł, zgodności z regułami biznesowymi, wyniku testów walidacyjnych oraz historii błędów dla podobnych przypadków.
- **Metryki dedykowane.** Klasyczne wskaźniki procesowe są uzupełniane miarami specyficznymi dla pracy modelu: trafność, jednostkowy koszt decyzji, czas na podjęcie decyzji, współczynnik akceptacji rekomendacji, wskaźnik korekt rekomendacji, współczynnik eskalacji, wskaźnik oparcia odpowiedzi na źródłach, odsetek spraw wymagających korekty.

2.3. ROZRÓŻNIENIE POJĘĆ: CZYM AI-FIRST NIE JEST

W praktyce pojęcie AI-First bywa używane w sposób rozszerzony, co prowadzi do nieporozumień. W niniejszym opracowaniu przyjęto następujące rozgraniczenia:

- **AI-First nie jest synonimem pełnej automatyzacji.** W procesie AI-First mogą – i z reguły powinny – istnieć etapy wykonywane przez człowieka. Różnica polega na uzasadnieniu obecności człowieka, nie na jej braku.
- **AI-First nie jest procesem AI-Ready z większą liczbą agentów.** Architektura AI-First wymaga przeprojektowania, a nie skalowania komponentów AI w istniejącej strukturze.
- **AI-First nie jest deklaracją kulturową.** Strategiczna deklaracja typu „jesteśmy organizacją AI-First" nie jest tożsama z istnieniem procesów AI-First. To pierwsze jest aspiracją; to drugie – konkretnym projektem procesu, modelem odpowiedzialności i zestawem mechanizmów kontrolnych.
- **AI-First nie wymaga zaprojektowania procesu od podstaw.** Możliwy jest jako głęboka rekonstrukcja procesu istniejącego, przy zachowaniu ciągłości operacyjnej i stopniowej migracji wolumenu spraw.

2.4. POJĘCIA TECHNICZNE: MODEL, AGENT, SYSTEM AI, AUTOMATYZACJA

Dla zachowania precyzji w opracowaniu rozróżniono następujące pojęcia:

- **Model (AI).** Wytrenowany komponent obliczeniowy generujący wynik na podstawie danych wejściowych i parametrów wyuczonych w procesie treningu. Współczesne duże modele językowe są modelami ogólnego przeznaczenia, zdolnymi obsługiwać wiele typów zadań. Sam model nie zawiera funkcji celu procesu, odpowiedzialności ani trwałej pamięci operacyjnej procesu – stan, pamięć, narzędzia oraz polityki działania są dostarczane przez aplikację, orkiestratora lub architekturę agentową.
- **Aplikacja AI.** Konkretne rozwiązanie biznesowe wykorzystujące jeden lub więcej modeli, opakowane warstwą interfejsu, integracji i polityk. Aplikacja AI może być pojedynczym narzędziem (np. asystent w przeglądarce) lub komponentem szerszego systemu.
- **Agent.** Komponent zdolny do autonomicznego planowania i realizacji wielokrokowych zadań w środowisku z dostępem do narzędzi (wywoływanie funkcji API, wyszukiwanie, działanie na plikach), działający w granicach polityk, uprawnień i środowiska wykonawczego ustanowionych przez projektanta. Agent charakteryzuje się stanem roboczym, podejmuje decyzje o sekwencji działań oraz może modyfikować przebieg pracy na podstawie wyników cząstkowych.
- **System wieloagentowy.** Architektura, w której wiele wyspecjalizowanych agentów współpracuje ze sobą w celu realizacji określonego celu – każdy z odrębną rolą, narzędziami i kontekstem, zarządzany przez orkiestrator.
- **Przepływ pracy i orkiestracja.** Wykonywalny opis sekwencji kroków procesu, w którym komponenty AI i klasyczne łączone są w deterministyczny przepływ. W odróżnieniu od agenta, przepływ pracy nie planuje samodzielnie kolejnych kroków. Sam przepływ pracy nie jest komponentem AI – jest mechanizmem orkiestrującym, który może łączyć komponenty AI, reguły deterministyczne i integracje systemowe.
- **Klasyczna automatyzacja.** Reguły, integracje i przepływy wykonywane bez udziału modeli AI – robotyzacja procesów (RPA), silniki reguł biznesowych (BRMS), deterministyczne przepływy

pracy. W procesach AI-First i AI-Ready współistnieją z komponentami AI jako uzupełnienie deterministyczne.

Rozróżnienia te mają znaczenie nie tylko terminologiczne. Wybór między modelem, agentem a systemem wieloagentowym ma bezpośrednie konsekwencje dla profilu ryzyka, kosztu wdrożenia, sposobu nadzoru i mechanizmów audytu. Jest decyzją architektoniczną, a nie technologicznym detalem.

3. ANATOMIA PORÓWNAWCZA PARADYGMATÓW PROCESOWYCH

Rzetelne porównanie obu paradygmatów wymaga zastosowania ram obejmujących osiem wymiarów analizy. Każdy z nich opisuje istotny aspekt projektowy, w którym AI-Ready i AI-First różnią się w sposób jakościowy, a nie ilościowy.

Wymiar	Proces AI-Ready	Proces AI-First
Punkt wyjścia projektowania	Istniejąca, działająca mapa procesu (As-Is). Adaptacja zachowuje zasadniczą sekwencję czynności i podział ról	Projektowanie od podstaw lub głęboka rekonstrukcja procesu istniejącego – bez założenia o sekwencji wynikającej z dotychczasowego sposobu pracy. Punktem wyjścia są: oczekiwany wynik, dane wejściowe, ograniczenia regulacyjne oraz ekonomika procesu
Domyślny wykonawca pracy poznawczej	Człowiek wspierany przez AI (asystent, narzędzie wspomagające, automatyzacja czynności rutynowych). AI pełni rolę narzędzia w rękach analityka lub specjalisty	Komponent AI – model, agent, klasyfikator, system wieloagentowy – jako domyślny wybór, o ile nie wykluczają tego ryzyko, regulacje, koszt błędu lub potrzeba eksperckiego osądu
Architektura informacyjna	Dane procesowe są opisane na tyle szczegółowo, by AI mogła z nich korzystać – ustrukturyzowane dane wejściowe, słowniki biznesowe, polityki i reguły	Architektura zbudowana wokół kontekstu modelu: warstwa wyszukiwania i pobierania kontekstu (RAG – *Retrieval-Augmented Generation*, GraphRAG), kontrolowany stan roboczy i pamięć operacyjna z polityką retencji i wersjonowania, wywoływanie narzędzi zewnętrznych, telemetria
Mechanizm decyzyjny	Reguły biznesowe pozostają wiążące: AI rekomenduje, człowiek zatwierdza decyzje o znaczącym wpływie. Decyzje o znaczącym wpływie biznesowym podlegają proporcjonalnemu mechanizmowi nadzoru człowieka	Decyzje rutynowe podejmowane są autonomicznie według polityki opartej na wielokryteriowej ocenie sygnałów jakości i ryzyka. Decyzje o znaczącym wpływie biznesowym są kierowane do recenzji z udokumentowaną odpowiedzialnością decyzyjną i możliwością zakwestionowania wyniku
Obsługa wyjątków	Wyjątki obsługuje człowiek według dotychczasowych ścieżek. AI sygnalizuje anomalie, ale nie podejmuje samodzielnie ich rozwiązania	Wyjątki klasyfikowane są przez agenta nadzorującego; rutynowe są rozwiązywane autonomicznie, nietypowe trafiają do recenzenta, operatora lub właściciela decyzji wraz z pełnym kontekstem i historią
Skalowalność i koszt jednostkowy	Koszt jednostkowy obsługi maleje liniowo lub schodkowo wraz ze wzrostem dojrzałości wdrożeń AI. Wąskim gardłem przepustowości pozostaje czas pracy operatorów ludzkich	Koszt jednostkowy może się zmniejszyć radykalnie, lecz nie redukuje się do kosztu wywołania modelu. Realny całkowity koszt jednostkowy obejmuje orkiestrację, pobieranie kontekstu, walidację, monitorowanie, logowanie, obsługę wyjątków, nadzór człowieka i utrzymanie bazy wiedzy

Mierniki efektywności	Klasyczne wskaźniki procesowe (czas cyklu, koszt jednostkowy, jakość) uzupełnione miarami adopcji AI, czasu zaoszczędzonego oraz akceptacji rekomendacji przez użytkowników	Nowe metryki dedykowane: trafność modelu, współczynnik akceptacji rekomendacji, wskaźnik korekt rekomendacji, współczynnik eskalacji, wskaźnik oparcia odpowiedzi na źródłach, koszt na decyzję, czas do decyzji, odsetek spraw wymagających korekty
Ład zarządczy i zgodność regulacyjna	Rozszerzenie istniejących polityk procesowych o klauzule dotyczące użycia AI, ochrony danych i ścieżek odwoławczych	Dedykowane ramy zarządzania AI: rejestr modeli i systemów AI, polityka danych, ramy oceny ryzyka modelowego, procedury incydentowe i mechanizmy odwoławcze – osadzone w systemie zarządzania AI (ISO/IEC 42001, AI Act, NIST AI RMF)

3.1. INTERPRETACJA WYBRANYCH WYMIARÓW

ARCHITEKTURA INFORMACYJNA

Znacząca część niepowodzeń wdrożeń AI wynika nie z ograniczeń modeli, lecz z niedostatecznej jakości warstwy informacyjnej. W procesie AI-Ready warstwa ta polega głównie na uporządkowaniu istniejących źródeł – dokumentów polityk, słowników biznesowych, baz przypadków. W procesie AI-First warstwa ta jest projektowana jako autonomiczna domena: zawiera wektorowe reprezentacje wiedzy, indeksy hybrydowe, mechanizmy aktualności danych, hierarchię źródeł według ich wiarygodności i mocy obowiązywania, zasady wersjonowania oraz kontrolowany stan roboczy i pamięć operacyjną agentów z polityką retencji, uprawnień i usuwania danych.

MECHANIZM DECYZYJNY I NADZÓR CZŁOWIEKA

W procesie AI-Ready wiążące pozostają reguły biznesowe: AI rekomenduje, a człowiek zatwierdza decyzje o znaczącym wpływie. Pozwala to zachować ciągłość procesu wobec audytu i regulatora. W procesie AI-First decyzje rutynowe są podejmowane autonomicznie, natomiast decyzje o wysokim wpływie podlegają adekwatnemu mechanizmowi nadzoru człowieka, możliwości zakwestionowania wyniku oraz udokumentowanej odpowiedzialności decyzyjnej. Wymaga to, aby reguły biznesowe były dostępne nie tylko jako tekst, lecz również w formie wykonywalnej (np. polityki w formie kodu). Szczegółowa typologia nadzoru została przedstawiona w rozdziale 5.

ESKALACJA: WIELOKRYTERIOWA OCENA SYGNAŁÓW ZAMIAST POJEDYNCZEGO PROGU

Eskalacja w procesach AI-First nie powinna być oparta wyłącznie na wskaźniku ufności zwracanym przez model. Modele językowe nie dostarczają natywnie precyzyjnie skalibrowanego wskaźnika ufności, a sama wartość tego wskaźnika – o ile system w ogóle ją udostępnia – nie jest wystarczającą podstawą decyzji o przekazaniu sprawy do recenzji człowieka. W praktyce próg eskalacji powinien być wyliczany na podstawie kilku sygnałów: klasy ryzyka sprawy, kompletności danych wejściowych, jakości i aktualności źródeł, zgodności odpowiedzi z regułami biznesowymi, wyniku testów walidacyjnych, liczby wykrytych anomalii oraz historii błędów dla podobnych przypadków. Takie podejście jest spójne

z duchem ram NIST AI Risk Management Framework, w których zarządzanie ryzykiem opiera się na funkcjach *Govern, Map, Measure* i *Manage*, a nie na pojedynczym wskaźniku.

SKALOWALNOŚĆ I KOSZT JEDNOSTKOWY

Najbardziej widoczna ekonomicznie różnica między paradygmatami dotyczy struktury kosztów. W procesie AI-Ready koszt jednostkowy obsługi maleje stopniowo wraz z dojrzałością wdrożeń AI, ale jego dolnym ograniczeniem pozostaje koszt pracy człowieka. W procesie AI-First koszt jednostkowy może spaść radykalnie w porównaniu z procesem opartym głównie na pracy ludzkiej, ale nie redukuje się wyłącznie do kosztu wywołania modelu.

Realny całkowity koszt jednostkowy w procesie AI-First obejmuje koszt orkiestracji, pobierania kontekstu, walidacji, monitorowania, logowania, obsługi wyjątków, nadzoru człowieka, utrzymania bazy wiedzy oraz ryzyka błędnych decyzji. Dlatego ekonomika AI-First powinna być liczona jako **całkowity koszt podjęcia decyzji lub obsługi sprawy, a nie koszt pojedynczego wywołania modelu**. Ta zmiana modelu kosztowego generuje wymierne korzyści, lecz jednocześnie zwiększa ryzyko systemowe: pojedynczy błąd projektowy w architekturze procesu może zostać powielony w dużym wolumenie spraw.

WYSZUKIWANIE I POBIERANIE KONTEKSTU (RAG): ZAKOTWICZENIE NIE JEST GWARANCJĄ PRAWDY

Architektura RAG istotnie zwiększa szansę na odpowiedź ugruntowaną w wiarygodnych źródłach, lecz nie eliminuje ryzyka błędów merytorycznych. Ryzyko to może powstać na etapie indeksowania, wyszukiwania, oceny trafności dokumentów, interpretacji fragmentów, rozstrzygania konfliktów między źródłami albo użycia nieaktualnej wersji procedury. Dlatego architektura RAG wymaga stosowania rozdzielonych warstw metryk:

- **metryki pobierania kontekstu:** precyzja doboru źródeł (precision@k), kompletność doboru (recall@k), pozycja w rankingu trafnych źródeł (MRR – *Mean Reciprocal Rank*, NDCG – *Normalized Discounted Cumulative Gain*), aktualność źródeł (*freshness*);
- **metryki odpowiedzi:** zgodność odpowiedzi ze źródłami, trafność odpowiedzi, stopień pokrycia odpowiedzi cytowaniami oraz poprawność cytowań;
- **metryki bezpieczeństwa:** trafność odmów odpowiedzi, wskaźnik odpowiedzi nieuzasadnionych dowodami oraz wskaźnik luk w wiedzy;
- **metryki procesu:** współczynnik eskalacji, wskaźnik korekt rekomendacji, wskaźnik ponownej obróbki sprawy, czas do decyzji.

Niezbędnym uzupełnieniem jest mechanizm odmowy odpowiedzi przy braku wystarczającego poparcia w źródłach – często ważniejszy niż sama jakość generowania. Hierarchia źródeł powinna być zaprojektowana w sposób umożliwiający rozróżnienie: źródła wiążące (regulaminy, polityki autoryzowane), referencyjne (dokumentacja procesowa), pomocnicze (przykłady, precedensy) oraz historyczne (archiwum decyzji).

4. MACIERZ DECYZYJNA: KIEDY STOSOWAĆ DANY PARADYGMAT

Wybór paradygmatu nie jest decyzją technologiczną, lecz projektową, podejmowaną na poziomie właściciela procesu i sponsora biznesowego. Poniższa macierz przedstawia sześć kryteriów decyzyjnych wraz ze wskazaniem paradygmatu, do którego naturalnie ciążą procesy o określonej charakterrystyce.

Kryterium decyzyjne	Wskazanie: AI-Ready	Wskazanie: AI-First
Wolumen i powtarzalność	Niski lub średni wolumen, wysoka różnorodność przypadków, znaczący udział wiedzy ukrytej	Wysoki wolumen i powtarzalność lub wysoki koszt jednostkowy, możliwa formalizacja kontekstu i kryteriów decyzji
Profil ryzyka decyzji	Decyzje regulowane, nieodwracalne lub o dużym wpływie na klienta wymagają co najmniej silnego nadzoru człowieka i rygorystycznej audytowalności (np. kredytowe, medyczne, sądowe, dostęp do usług podstawowych); zwykle przemawia to za wskazaniem AI-Ready albo ograniczonego wariantu AI-First z trybem człowiek w pętli	Decyzje rutynowe, odwracalne, z jasnymi kryteriami akceptacji i niskim kosztem korekty błędu
Dojrzałość danych i wiedzy	Dane są rozproszone, jakościowo niejednorodne, wiedza ekspercka pozostaje w głowach pracowników, brak ustrukturyzowanej bazy wiedzy	Dostępna ustrukturyzowana baza wiedzy, dokumentacja procesowa, dane historyczne pozwalające na ewaluację jakości decyzji
Wymagania regulacyjne	Klasyfikacja jako system wysokiego ryzyka w rozumieniu AI Act lub silne wymogi sektorowe – audytowalność, wyjaśnialność, dokumentacja, mechanizmy odwoławcze i adekwatny nadzór człowieka	Wymogi proporcjonalne i spełnialne przez logowanie wywołań, polityki retencji, wersjonowanie modeli i kontrolowane bramy decyzyjne
Tolerancja zmiany organizacyjnej	Niska – silne przywiązanie do istniejących ról, niska gotowość do redefinicji modelu odpowiedzialności i podziału pracy	Wysoka – sponsor zarządczy gotowy na przeprojektowanie podziału pracy, ról stanowiskowych i modelu odpowiedzialności
Ekonomika i horyzont czasowy	Potrzeba szybkich rezultatów (3-9 miesięcy), ograniczony budżet inwestycyjny, presja na zwrot z inwestycji w krótkim terminie	Horyzont średnioterminowy (12-24 miesiące), budżet inwestycyjny pozwalający na fazę projektowania, walidacji i pilotażu, dodatnia oczekiwana wartość NPV

4.1. MODEL EKONOMICZNY ZAMIAST PROGU WOLUMENU

Często stawiane pytanie „jaki wolumen spraw uzasadnia stosowanie podejścia AI-First" jest źle postawione. Procesy o wysokim koszcie jednostkowym mogą uzasadniać AI-First nawet przy niskim wolumenie. Natomiast procesy zdominowane przez tanią pracę rutynową często podważają ekonomiczny sens transformacji – i to niezależnie od skali wolumetrycznej. Właściwym narzędziem decyzyjnym jest **reguła ekonomiczna AI-First**.

Reguła ekonomiczna AI-First

NPV AI-First = (wolumen × oszczędność jednostkowa lub dodatkowa wartość na sprawę) – (koszt wdrożenia + koszt operacyjny + koszt nadzoru + koszt zgodności + oczekiwana strata z błędów) – w analizowanym horyzoncie czasowym i przy uzgodnionej stopie dyskontowej.

Decyzja o przejściu do AI-First powinna być podjęta wówczas, gdy wartość NPV przedsięwzięcia jest dodatnia po uwzględnieniu odpowiedniego bufora niepewności, a profil ryzyka pozwala na spełnienie wymogów regulacyjnych. Kalkulacja powinna obejmować nie tylko koszt licencji modelu, lecz wszystkie składowe wymienione w rozdziale 3.1. (akapit o skalowalności).

4.2. WSTĘPNY FILTR DECYZYJNY

Zespoły potrzebujące szybkiego rozeznania mogą zastosować filtr wstępny pozwalający uporządkować dyskusję na temat gotowości do uruchomienia ścieżki rekonstrukcyjnej. Filtr nie zastępuje pełnej analizy ekonomicznej i regulacyjnej: służy do wczesnego zidentyfikowania luk wymagających zamknięcia przed podjęciem decyzji o przejściu do paradygmatu AI-First.

Na filtr wstępny składają się odpowiedzi na następujące pytania:

1. **Czy prognozowany zwrot z inwestycji jest dodatni?** Innymi słowy: czy oszczędności wynikające z efektu skali oraz redukcji kosztu jednostkowego rekompensują nakłady na wdrożenie, koszty operacyjne, koszty nadzoru, koszty zgodności regulacyjnej oraz oczekiwaną stratę z tytułu błędnych decyzji zautomatyzowanych?
2. **Czy organizacja dysponuje wersjonowaną, oficjalną bazą wiedzy** oraz danymi historycznymi pozwalającymi na ewaluację jakości decyzji modelu?
3. **Czy klasyfikacja systemu AI lub konkretnego przypadku użycia w rozumieniu AI Act pozwala na przyjęcie zakładanego poziomu autonomii w paradygmacie AI-First?** Procesy klasyfikowane jako wysokiego ryzyka mogą wymagać silniejszych mechanizmów nadzoru, trudnych do pogodzenia z autonomią agentową.
4. **Czy koszt korekty błędnej decyzji jest niski** lub czy istnieje mechanizm wczesnego wykrywania błędów przed ich materializacją?
5. **Czy istnieje sponsor zarządczy gotowy na przeprojektowanie ról i odpowiedzialności?** Bez mandatu zarządczego program AI-First z reguły zatrzymuje się na poziomie pilotażu.

6. **Czy organizacja dysponuje kompetencjami do utrzymania architektury agentowej?** Zespół do projektowania, monitorowania, testowania i obsługi incydentów – własny lub partnerski.
7. **Czy istnieje referencyjny zbiór testów regresyjnych pozwalający na ciągłą ocenę jakości?** Bez obiektywnego zbioru testów każda zmiana modelu lub instrukcji opiera się na niezweryfikowanym założeniu jakościowym.

Pozytywna odpowiedź na wszystkie siedem pytań uzasadnia rozpoczęcie projektu w paradygmacie AI-First. **Negatywna odpowiedź na jedno lub więcej pytań nie wyklucza ścieżki rekonstrukcyjnej, lecz wskazuje konieczność zastosowania działań kompensacyjnych**: zawężenia zakresu pilotażu, wzmocnienia trybu nadzoru człowieka, uzupełnienia bazy wiedzy, ograniczenia autonomii agenta, doinwestowania kompetencji albo czasowego pozostania na etapie AI-Ready do momentu zamknięcia luki. Dla każdej luki należy określić plan działań naprawczych – luka nie może być ignorowana czy przemilczana.

5. TYPOLOGIA NADZORU CZŁOWIEKA I TYPÓW INTERAKCJI CZŁOWIEK-AI

Schemat „AI rekomenduje, człowiek zatwierdza" jest uproszczeniem użytecznym, lecz niewystarczającym dla projektowania procesów AI-First i AI-Ready. W praktyce projektowania systemów AI stosuje się kilka wzorców nadzoru, różniących się stopniem zaangażowania człowieka oraz typem ryzyka, jakim zarządzają. Wzorce te nie tworzą jednolicie zdefiniowanego standardu prawnego – funkcjonują jako użyteczna typologia projektowa. W zakresie nadzoru człowieka AI Act wymaga, aby nadzór nad systemami wysokiego ryzyka był skuteczny, proporcjonalny do ryzyka, poziomu autonomii i kontekstu użycia oraz umożliwiał osobom nadzorującym monitorowanie działania systemu, interpretowanie jego wyników, odrzucenie wyniku, uchylenie lub odwrócenie skutku decyzji oraz zatrzymanie działania systemu. W niniejszym opracowaniu zastosowano cztery wzorce, dla których w polskiej literaturze zaczynają się utrwalać następujące odpowiedniki: człowiek w pętli decyzyjnej (*human-in-the-loop*), człowiek nadzorujący pętlę (*human-on-the-loop*), człowiek zarządzający pętlą decyzyjną (*human-over-the-loop*) oraz tryb bez udziału człowieka w pojedynczej decyzji (*human-out-of-the-loop*).

Należy rozróżnić **dwa poziomy opisu współpracy człowieka z komponentem AI**. Pierwszy poziom to **tryb nadzoru człowieka nad decyzją lub akcją systemową**. Odpowiada on na pytanie, czy człowiek zatwierdza wynik przed wywo-łaniem skutku, monitoruje działanie systemu, zarządza mechanizmem decyzyjnym na poziomie polityk i metryk, czy też nie uczestniczy w pojedynczej decyzji. Drugi poziom to **typ interakcji człowiek-AI w etapie procesu**. Odpowiada on na pytanie, jak człowiek i agent współdziałają w danym fragmencie procesu: autonomicznie, pod nadzorem, przez zatwierdzenie, równolegle, przez eskalację albo w pętli informacji zwrotnej.

Tryb nadzoru	Charakterystyka	Typowe zastosowanie
Człowiek w pętli decyzyjnej (*human-in-the-loop*)	Człowiek dokonuje oceny, akceptacji, odrzucenia albo modyfikacji wyniku systemu przed wywołaniem skutku decyzji. Ręczna walidacja dotyczy decyzji z określonej klasy ryzyka, decyzji przekraczających ustalone progi lub decyzji wywołujących istotny skutek prawny, finansowy, medyczny, organizacyjny albo społeczny	Decyzje wysokiego wpływu i wymagające uzasadnialności, kontroli oraz możliwości zakwestionowania wyniku: decyzje kredytowe, medyczne, prawne, dyscyplinarne, kadrowe, ubezpieczeniowe, dotyczące dostępu do usług podstawowych lub świadczeń publicznych
Człowiek nadzorujący pętlę (*human-on-the-loop*)	System wykonuje decyzje lub rekomendacje autonomicznie w określonych granicach, a człowiek monitoruje jego działanie, wyniki, alerty, wyjątki i anomalie. Ma uprawnienia do interwencji, korekty, eskalacji, zatrzymania działania systemu albo wycofania skutku decyzji	Procesy rutynowe, powtarzalne, o umiarkowanym ryzyku, w których potrzebna jest szybka reakcja na anomalie, błędy, dryf modelu lub przekroczenie progów jakości: obsługa zgłoszeń, monitoring jakości, klasyfikacja spraw, automatyczna priorytetyzacja, wstępna detekcja nieprawidłowości
Człowiek zarządzający pętlą (*human-over-the-loop*)	Człowiek nie kontroluje pojedynczych decyzji operacyjnych, lecz zarządza całym mechanizmem decyzyjnym: politykami, regułami, progami ryzyka, testami ewaluacyjnymi, metrykami	Procesy o stabilnej jakości, dobrze opisanych regułach, mierzalnych parametrach działania i niskim lub umiarkowanym koszcie pojedynczego błędu, w których większą wartość daje

	jakości, wyjątkami, audytem, kalibracją, monitoringiem dryfu, zatwierdzaniem zmian i obsługą incydentów	nadzór systemowy niż walidacja każdej decyzji
Bez udziału człowieka w pojedynczej decyzji (*human-out-of-the-loop*)	System podejmuje i wykonuje pojedyncze decyzje bez bieżącego udziału człowieka. Udział człowieka pozostaje na poziomie projektowania, zatwierdzania zasad działania, oceny ryzyka, monitoringu ex post, audytu, obsługi incydentów i okresowej rekonfiguracji systemu.	Decyzje niskiego ryzyka, odwracalne, o niskim koszcie błędu, nieingerujące istotnie w prawa lub sytuację osoby, objęte silnymi mechanizmami kontroli, logowania, progami bezpieczeństwa i formalną oceną ryzyka przed uruchomieniem.

Tabela poniżej mapuje sześć typów interakcji na cztery tryby nadzoru człowieka opisane w powyższej tabeli opisującej tryby nadzoru. Mapowanie ma charakter projektowy, a nie regulacyjny.

Symbol interkacji	Typ interakcji	Definicja operacyjna	Relacja do trybu nadzoru	Komentarz projektowy
A	Autonomiczny	Agent wykonuje czynność lub akcję systemową bez bieżącej ingerencji człowieka	Najbliższy odpowiednik: *human-out-of-the-loop*	Dopuszczalne przede wszystkim dla decyzji niskiego ryzyka, odwracalnych albo objętych silną kontrolą ex post
N	Nadzorowany	Agent działa w określonych granicach, a człowiek monitoruje wyniki, alerty, wyjątki i anomalie	Odpowiada *human-on-the-loop*	Właściwy dla procesów rutynowych, w których potrzebna jest szybka interwencja po wykryciu anomalii, dryfu jakości lub przekroczenia progu
Z	Zatwierdzenie	Agent przygotowuje wynik, rekomendację, decyzję lub treść, a człowiek akceptuje, odrzuca albo modyfikuje wynik przed wywołaniem skutku	Odpowiada *human-in-the-loop*	Właściwy dla decyzji wysokiego wpływu, decyzji o skutku finansowym, prawnym, regulacyjnym, medycznym, kadrowym lub społecznym
R	Równoległy	Człowiek i agent pracują równocześnie nad tym samym zadaniem, dzieląc się rolami	Brak bezpośredniego odpowiednika w typologii nadzoru	To wzorzec koprodukcji wyniku, a nie tryb kontroli. Dla decyzji kończącej etap należy osobno wskazać tryb nadzoru
E	Eskalacja	Agent przekazuje sprawę człowiekowi po spełnieniu kryteriów ryzyka, niejednoznaczności, braku danych, anomalii lub przekroczenia progu decyzyjnego	Nie jest trybem nadzoru. Jest mechanizmem przełączenia do *human-in-the-loop* albo *human-on-the-loop*	W raporcie należy traktować eskalację jako politykę przełączenia kontroli, a nie jako samodzielny poziom autonomii
F	Feedback	Człowiek ocenia wynik pracy agenta, oznacza	Nie jest trybem nadzoru nad pojedynczą decyzją.	Feedback zasila kalibrację, monitoring

		błędy, wskazuje potrzebę korekty reguł, instrukcji, bazy wiedzy albo modelu	Może wspierać *human-over-the-loop*	jakości, audyt, uczenie organizacyjne i governance. Nie powinien samodzielnie opisywać nadzoru nad skutkiem decyzji

Wybór trybu nadzoru **nie jest decyzją statyczną**. W ramach jednego procesu AI-First lub AI-Ready mogą współistnieć różne tryby nadzoru – przypisane do różnych decyzji, akcji systemowych, kategorii spraw lub poziomów ryzyka. Niezależnie od tego każdy etap procesu może mieć odrębny typ interakcji człowiek-AI. Typ interakcji nie przesądza automatycznie o trybie nad-zoru: etap oznaczony jako R (Równoległy) może zakończyć się decyzją wymagającą zatwierdzenia, etap E (Eskalacja) może przełączać sprawę do recenzenta, a etap F (Feedback) może dotyczyć wyłącz-nie uczenia organizacyjnego i nie obejmować decyzji wobec klienta.

Kluczowe jest świadome przypisanie obu kategorii: trybu nadzoru do decyzji, rekomendacji lub akcji systemowej oraz typu interakcji do etapu pracy. Ocena zdolności kredytowej, scoring klienta lub decyzja kredytowa dotycząca osoby fizycznej może wymagać silnego nadzoru człowieka zgodnie z klasyfikacją ryzyka systemu, wymogami AI Act, RODO oraz regulacjami sektorowymi, podczas gdy generowanie rutynowej korespondencji wewnętrznej może być realizowane bez udziału człowieka w pojedynczej decyzji. W obu przypadkach powinien opisywać sposób współpracy operacyjnej, ale nie zastępuje analizy nadzoru regulacyjnego.

AI Act wymaga, by nadzór człowieka nad systemami wysokiego ryzyka był **adekwatny do ryzyka, poziomu autonomii i kontekstu użycia**, a osoby nadzorujące posiadały kompetencje, uprawnienia i wsparcie wystarczające do zrozumienia działania systemu, monitorowania jego pracy, interpretowania wyników i interwencji w razie potrzeby. Dla systemów wysokiego ryzyka niedopuszczalny jest brak realnego, skutecznego i udokumentowanego nadzoru człowieka. Zakres udziału człowieka w pojedynczej decyzji powinien wynikać z oceny ryzyka, przeznaczenia systemu, przepisów sektorowych oraz wymagań AI Act i RODO.

6. ŚCIEŻKI TRANSFORMACJI

Aby uniknąć błędnej interpretacji obu paradygmatów jako kolejnych poziomów tej samej skali dojrzałości, w niniejszym opracowaniu wyróżniono dwie odrębne ścieżki transformacji procesowej. Pierwsza z nich – ścieżka adaptacyjna – prowadzi od procesu klasycznego przez stopniowe wsparcie AI do procesu AI-Ready. Druga – ścieżka rekonstrukcyjna – jest uruchamiana zwykle po osiągnięciu dojrzałości AI-Ready w procesie istniejącym i stanowi decyzję architektoniczną o przeprojektowaniu procesu, nie zaś o jego dalszym skalowaniu. W przypadku procesów projektowanych od podstaw paradygmat AI-First może być zastosowany od początku – pod warunkiem równoczesnego wbudowania standardów AI-Ready.

Kluczowe rozróżnienie. Należy podkreślić krytyczne założenie koncepcyjne: paradygmat AI-First nie stanowi „etapu czwartego" na skali dojrzałości procesu po osiągnięciu poziomu AI-Ready. Jest odrębnym przedsięwzięciem projektowym, które organizacja rozpoczyna wtedy, gdy osiągnie wystarczającą gotowość informacyjną, regulacyjną i organizacyjną oraz dostatecznie mocne uzasadnienie ekonomiczne poniesienia kosztu rekonstrukcji procesu.

Etap 1 As-Is ***Proces klasyczny***	**Etap 2** **AI-Augmented** ***Wsparcie punktowe***	**Etap 3 AI-Ready** ***Proces zaadaptowany***	**Ścieżka odrębna:** **AI-First** ***Decyzja architektoniczna***
Cała praca poznawcza wykonywana przez człowieka. Brak komponentów AI lub punktowe nieautoryzowane użycie AI przez pracowników (niezgodnie z polityką organizacji)	Pojedyncze czynności wspierane przez AI (streszczanie, klasyfikacja, generowanie szkiców). Proces nie został jeszcze przeprojektowany	Proces przygotowany do współpracy człowiek-AI: opisane dane wejściowe, kryteria jakości, polityki, ścieżki eskalacji, audytowalność	Decyzja o rekonstrukcji procesu – a nie kolejny poziom dojrzałości. Komponent AI jako domyślny kandydat do wykonania pracy poznawczej. Człowiek pełni role projektanta, recenzenta, arbitra eskalacji i właściciela ryzyka

6.1. ŚCIEŻKA ADAPTACYJNA: AS-IS → AI-AUGMENTED → AI-READY

Ścieżka adaptacyjna jest właściwa dla zdecydowanej większości procesów w polskich organizacjach. Polega ona na sekwencyjnym przygotowywania procesu do współpracy z komponentami AI, przy zachowaniu zasadniczej logiki biznesowej i stopniowym wprowadzaniu nowych ról kontrolnych, artefaktów informacyjnych oraz mechanizmów audytu.

BRAMA 1: AS-IS → AI-AUGMENTED

Warunki przejścia:

- zidentyfikowane miejsca w procesie, w których wsparcie AI przyniesie mierzalną wartość;
- przyjęta polityka korzystania z AI (akceptowalne kategorie danych, dopuszczone narzędzia, sposoby weryfikacji wyników) – eliminująca nieautoryzowane użycie AI przez pracowników;

- podstawowe szkolenie zespołu w projektowaniu zapytań i instrukcji dla modelu (inżynieria zapytań) oraz krytycznej oceny generowanych odpowiedzi – w ramach kompetencji AI (*AI literacy*) wymaganych przez AI Act.

BRAMA 2: AI-AUGMENTED → AI-READY

Warunki przejścia:

- opisane dane wejściowe, dane wyjściowe oraz kryteria akceptacji dla każdej czynności wspieranej przez AI;
- uporządkowana baza wiedzy z mechanizmem aktualizacji, hierarchią źródeł (źródła wiążące, referencyjne, pomocnicze, historyczne) i wyznaczonym właścicielem treści;
- zdefiniowane ścieżki eskalacji oraz polityka eskalacji oparta na wielokryteriowej ocenie sygnałów jakości i ryzyka;
- wdrożone logowanie wywołań i podstawowe metryki jakości obejmujące co najmniej współczynnik akceptacji rekomendacji, wskaźnik korekt oraz wskaźnik oparcia odpowiedzi na źródłach.

6.2. ŚCIEŻKA REKONSTRUKCYJNA: AI-READY → DECYZJA ARCHITEKTONICZNA → AI-FIRST

Ścieżka rekonstrukcyjna nie jest naturalną kontynuacją ścieżki adaptacyjnej. Dla procesów istniejących stanowi odrębną decyzję projektową, podejmowaną zwykle wówczas, gdy proces AI-Ready działa stabilnie w produkcji, organizacja dysponuje wymaganą dojrzałością informacyjną i regulacyjną, a wynik ekonomiczny głębokiej rekonstrukcji jest dodatni. W przypadku procesów projektowanych od podstaw decyzja o paradygmacie AI-First może być podjęta na etapie projektowania docelowego modelu operacyjnego. W obu przypadkach decyzja wymaga formalnego mandatu sponsora zarządczego oraz zatwierdzenia profilu ryzyka przez właściciela ryzyka modelowego.

BRAMA 3: WARUNKI URUCHOMIENIA ŚCIEŻKI REKONSTRUKCYJNEJ

Decyzja o uruchomieniu projektu AI-First powinna zostać podjęta po pozytywnej weryfikacji następujących warunków:

- dojrzała baza wiedzy o jakości potwierdzonej w produkcyjnej pracy z procesem AI-Ready;
- zestaw testów ewaluacyjnych i regresyjnych pozwalający na obiektywną ocenę zmian w modelu, instrukcji lub kontekście;
- dedykowane ramy zarządzania dla AI (*AI governance*): rejestr modeli i systemów AI, polityki danych, procedury incydentowe, mechanizmy odwoławcze;
- dodatnia oczekiwana wartość NPV przedsięwzięcia w analizowanym horyzoncie czasowym (zob. rozdział 4.1) z uwzględnieniem kosztu kapitału, kosztu zgodności i oczekiwanej straty z błędów;
- dopasowanie trybu nadzoru człowieka do klasyfikacji ryzyka systemu w rozumieniu AI Act;
- decyzja architektoniczna o przeprojektowaniu procesu (nie jego skalowaniu) wraz z formalnym mandatem sponsora zarządczego oraz potwierdzeniem ze strony właściciela ryzyka modelowego.

Niespełnienie któregokolwiek z powyższych warunków nie zamyka trwale drogi do ścieżki rekonstrukcyjnej. Oznacza jednak, że uruchomienie programu AI-First w aktualnych warunkach organizacyjnych generowałoby niedopuszczalny dług architektoniczny. Właściwym kierunkiem działania jest wówczas wstrzymanie migracji oraz konsekwentne domykanie zidentyfikowanych luk w warstwie AI-Ready, a nie przyspieszanie deklaracji wyprzedzających zdolności wykonawcze. **Próba uruchomienia ścieżki rekonstrukcyjnej z pominięciem któregoś z warunków generuje dług architektoniczny**, który materializuje się jako ryzyko operacyjne, regulacyjne lub finansowe w fazie produkcyjnej.

7. RYZYKA I MECHANIZMY ICH OGRANICZANIA

Niezależnie od wybranego paradygmatu wdrożenia AI w procesach niosą ze sobą określone ryzyka. Dla przejrzystości ryzyka podzielono na dwie grupy: ryzyka klasyczne wdrożeń modeli generatywnych oraz ryzyka bezpieczeństwa specyficzne dla architektur agentowych.

7.1. RYZYKA KLASYCZNE WDROŻEŃ AI

Ryzyko	Charakterystyka	Mechanizm ograniczania ryzyka
Halucynacje modelu	Generowanie odpowiedzi pozornie wiarygodnych, ale niezgodnych z faktami, nieopartych na wiarygodnym źródle	Ugruntowanie odpowiedzi w wiarygodnych źródłach (RAG) wraz z metrykami jakości wyszukiwania, walidacja przez niezależne reguły deterministyczne, źródła referencyjne, ewaluację ekspercką lub pomocniczy model oceniający, obowiązkowy nadzór człowieka dla decyzji o znaczącym wpływie biznesowym
Błędy wyszukiwania i niedostateczne oparcie odpowiedzi na źródłach (*grounding failure*)	RAG może pobrać niewłaściwy fragment, dokument nieaktualny, dokument o niższej randze lub pozostawić nierozwiązany konflikt między źródłami	Hierarchia wiarygodności i mocy obowiązywania dokumentów, wersjonowanie źródeł, metryki jakości wyszukiwania – precyzja doboru źródeł, pokrycie odpowiedzi cytowaniami, aktualność źródeł, wskaźnik luk w bazie wiedzy – oraz mechanizm odmowy odpowiedzi przy braku wystarczającej podstawy
Dryf modelu i regresja jakości	Zmiana zachowania modelu po aktualizacji wersji lub w wyniku zmieniającego się kontekstu danych	Wersjonowanie modeli i instrukcji dla modelu, zestaw testów ewaluacyjnych i regresyjnych, monitorowanie wskaźników jakości w produkcji, stopniowanie zakresu wdrożeń (tzw. wdrożenia kanarkowe) i procedury wycofania zmian
Niedostateczna wyjaśnialność	Trudność w uzasadnieniu decyzji modelu wobec klienta, audytora lub regulatora – kluczowa przy wymogach AI Act dla systemów wysokiego ryzyka	Wymóg generowania uzasadnień wraz z odpowiedzią, prowadzenie audytu wywołań, mapowanie decyzji na cytowane fragmenty bazy wiedzy, dokumentacja techniczna systemu
Stronniczość i nierówne traktowanie	Model lub reguły eskalacji mogą systematycznie pogarszać jakość obsługi określonych grup klientów, kategorii spraw lub kanałów kontaktu – szczególnie istotne w obszarach HR, kredytów, edukacji, usług publicznych oraz obsługi sporów	Testy bezstronności algorytmicznej (*fairness*) na zdefiniowanych segmentach, analiza błędów według grup, audyt danych treningowych i historycznych, kontrola cech wrażliwych i ich proxy w danych wejściowych, mechanizm odwoławczy dla osób objętych decyzją,

		monitorowanie odsetka decyzji negatywnych w segmentach.
Wyciek danych i ujawnienie informacji	Niezamierzone ujawnienie danych wrażliwych w instrukcjach, kontekście lub odpowiedziach modelu; nieautoryzowane wyprowadzenie danych przez wywoływane narzędzia	Klasyfikacja danych przed wprowadzeniem do modelu, polityka prywatności od fazy projektowania, wdrożenia lokalne dla kategorii danych wrażliwych, maskowanie, anonimizacja lub pseudonimizacja danych osobowych, kontrola uprawnień agentów do narzędzi i źródeł
Erozja kompetencji eksperckich	Spadek umiejętności pracowników wskutek przejęcia rutynowych czynności przez AI; utrata zdolności krytycznej oceny rekomendacji	Programy podtrzymujące kompetencje, obowiązkowe rotacje na zadaniach krytycznych, świadome projektowanie miejsc, w których człowiek wykonuje pracę samodzielnie, wymóg kompetencji AI (*AI literacy*) zgodnie z AI Act
Uzależnienie od dostawcy	Wysoki koszt zmiany modelu lub platformy, podatność na zmiany cennika i polityk dostawcy, zależność od formatów danych, instrukcji i mechanizmów orkiestracji	Architektura niezależna od modelu (warstwa abstrakcji), wybór otwartych standardów integracyjnych (Model Context Protocol, OpenAPI), strategia wielodostawcza dla zastosowań krytycznych, eksportowalność danych i konfiguracji

7.2. RYZYKA BEZPIECZEŃSTWA AGENTOWEGO

Architektury agentowe – kluczowe dla procesów AI-First – wprowadzają nową klasę ryzyk, która nie istnieje (lub istnieje w znacznie ograniczonym zakresie) w procesach klasycznych i AI-Ready. Ryzyka te wynikają z połączenia trzech właściwości: agent działa autonomicznie, ma dostęp do narzędzi z realnym wpływem na świat zewnętrzny oraz interpretuje dane z różnych źródeł, w tym takie, które mogą być spreparowane przez atakującego.

Wektor ryzyka	Charakterystyka	Mechanizm ograniczania ryzyka
Bezpośrednie wstrzyknięcie instrukcji (*prompt injection*)	Złośliwa instrukcja w danych wejściowych przejmuje kontrolę nad zachowaniem modelu i wymusza działania niezgodne z polityką	Walidacja, oczyszczanie i neutralizacja danych wejściowych, separacja warstwy instrukcji systemowej i danych użytkownika, polityki dopuszczalnych działań, monitorowanie odchyleń od oczekiwanego zachowania
Pośrednie wstrzyknięcie instrukcji (*indirect prompt injection*)	Złośliwa instrukcja osadzona w dokumencie, wiadomości e-mail lub stronie internetowej, samodzielnie pobrana przez agenta i zinterpretowana jako polecenie	Klasyfikacja źródeł według poziomu zaufania, obowiązkowe potwierdzenie przez człowieka działań o nieodwracalnych skutkach, izolacja kontekstu narzędzi od kontekstu modelu, audyt akcji agenta

Zatrucie bazy wiedzy (*knowledge base poisoning*)	Wprowadzenie do bazy wiedzy spreparowanych dokumentów, które wpływają na odpowiedzi modelu w sposób pożądany przez atakującego	Kontrola źródeł i procesu indeksowania, autoryzacja wprowadzania treści, hierarchia wiarygodności, regularne audyty zawartości bazy, mechanizmy detekcji anomalii
Nadmiarowe uprawnienia agenta	Agent ma szersze uprawnienia do narzędzi i danych, niż wymaga tego jego rola, co zwiększa skalę potencjalnej szkody	Zasada najmniejszych uprawnień, wąskie zakresy dostępu dla każdego narzędzia, dynamiczna autoryzacja akcji, segmentacja środowisk wykonawczych
Wyprowadzenie danych przez narzędzia (*data exfiltration*)	Atakujący wykorzystuje wywoływanie narzędzi (np. wysyłka e-maila, zapytanie do API zewnętrznego) do nieautoryzowanego wyprowadzenia danych	Lista dozwolonych adresatów i celów, klasyfikacja danych przed wywołaniem narzędzia, kontrola treści wychodzącej, audyt wszystkich wywołań zewnętrznych
Błędne lub nieodwracalne akcje przez API	Agent wykonuje działanie o nieodwracalnych skutkach (transakcja finansowa, usunięcie zasobu) na podstawie błędnej interpretacji instrukcji	Lista akcji wymagających potwierdzenia człowieka, środowiska odizolowane (*sandbox*) dla działań nowych, mechanizmy wycofania zmian, limity ilościowe i progi alarmowe
Kompromitacja łańcucha dostaw narzędzi	Złośliwy lub skompromitowany serwer narzędzi (np. publiczny serwer MCP), nadmiarowe uprawnienia w kluczach API, wyciek tokenów dostępowych z kontekstu agenta, podstawienie narzędzia o tej samej nazwie i innym działaniu	Lista zaufanych źródeł narzędzi z weryfikacją podpisów lub rejestru, zasada najmniejszych uprawnień przy wystawianiu kluczy, częsta rotacja sekretów i krótkie okresy ważności poświadczeń, izolacja sekretów od kontekstu modelu, monitoring i alertowanie nietypowych wywołań, weryfikacja tożsamości serwera narzędzi

Wymienione mechanizmy ograniczania ryzyka nie są opcjonalne. Dla procesów AI-First powinny zostać wdrożone **przed** uruchomieniem produkcyjnym; dla procesów AI-Ready – zsynchronizowane z dojrzewaniem wdrożeń komponentów AI. Pominięcie tych mechanizmów w środowisku produkcyjnym istotnie zwiększa ryzyko, że projekt generatywnej AI utknie w fazie pilotażu i nie osiągnie stabilnego wdrożenia produkcyjnego.

8. PRZYKŁAD PROCESU: OBSŁUGA REKLAMACJI KLIENTA

Aby osadzić ramy pojęciowe w konkretnym kontekście, poniżej przedstawiono proces obsługi reklamacji klienta w trzech wariantach: klasycznym, AI-Ready oraz AI-First. Przykład jest celowo uproszczony (pomija wiele uwarunkowań sektorowych i operacyjnych) lecz oddaje istotę różnic architektonicznych między paradygmatami.

Etap procesu	Klasyczny	AI-Ready	AI-First
Przyjęcie zgłoszenia	Pracownik czyta wiadomość klienta i ręcznie wprowadza dane do systemu CRM	AI wyodrębnia kluczowe dane (numer zamówienia, kategoria, intencja) i wstępnie wypełnia formularz; pracownik weryfikuje	Agent przyjmuje zgłoszenie z wielu kanałów (e-mail, czat, formularz), automatycznie klasyfikuje, wyodrębnia dane i tworzy sprawę
Klasyfikacja sprawy	Pracownik kategoryzuje reklamację na podstawie wewnętrznego słownika i własnego osądu	AI proponuje klasyfikację, pracownik akceptuje lub koryguje. Korekty są zapisywane jako dane zwrotne do okresowej ewaluacji jakości i kalibracji instrukcji lub reguł klasyfikatora	Agent wykonuje wieloetapową klasyfikację sprawy obejmującą kategorię, podkategorię, poziom ryzyka i priorytet, a także uwzględnia dodatkowe atrybuty klienta (np. status klienta strategicznego). Sprawy o niejednoznacznej klasyfikacji są kierowane do recenzenta według progu eskalacji opartego na wielokryteriowej ocenie sygnałów
Pobranie kontekstu i podstawy	Pracownik manualnie wyszukuje w kilku systemach regulamin, historię zamówień, korespondencję	Mechanizm wyszukiwania (RAG) pobiera odpowiedni fragment regulaminu i dane zamówienia; pracownik weryfikuje aktualność	Agent samodzielnie wywołuje narzędzia: bazę zamówień, regulamin, historię klienta, system płatności. Wyniki są oznaczone wersją źródła i datą
Decyzja o uznaniu reklamacji	Pracownik analizuje sprawę i podejmuje decyzję na podstawie polityk i własnego osądu	AI proponuje decyzję wraz z uzasadnieniem i cytatami z polityki; pracownik zatwierdza, koryguje lub odrzuca	Agent podejmuje decyzję autonomicznie dla spraw mieszczących się w polityce niskiego ryzyka. Sprawy przekraczające próg eskalacji oparty na wielokryteriowej ocenie sygnałów (klasa ryzyka, kompletność danych, jakość źródeł, wartość sprawy, historia klienta, anomalie, konflikty reguł) są kierowane do recenzji człowieka

Komunikacja z klientem	Pracownik samodzielnie redaguje odpowiedź według wzorca lub od podstaw	AI generuje szkic odpowiedzi dostosowany do kategorii sprawy; pracownik edytuje i wysyła	Agent wysyła odpowiedź automatycznie wyłącznie dla spraw prostych, niskowartościowych, odwracalnych i jednoznacznie mieszczących się w zatwierdzonej polityce niskiego ryzyka; pozostałe sprawy przekazuje do akceptacji recenzenta
Zamknięcie i archiwizacja	Pracownik ręcznie aktualizuje status sprawy i archiwizuje korespondencję	AI proponuje status i etykiety; archiwizacja półautomatyczna z weryfikacją	Agent zamyka sprawę, aktualizuje wskaźniki, generuje wpis audytowy z pełnym śladem wywołań modelu i decyzji

8.1. KOMENTARZ INTERPRETACYJNY

Z porównania wariantów wynika kilka obserwacji o znaczeniu praktycznym:

- **Wariant AI-Ready zwykle zachowuje zasadniczą sekwencję procesu i odpowiedzialność biznesową**, lecz wprowadza nowe role kontrolne (właściciel kontekstu, recenzent jakości odpowiedzi), nowe artefakty informacyjne, dodatkowe punkty walidacji oraz rejestr decyzji i uzasadnień. Nie jest to więc „ten sam proces z dodatkową funkcją AI", lecz proces zaadaptowany do współpracy z komponentami AI.
- **Wariant AI-First fundamentalnie reorganizuje podział ról**. Pracownik nie obsługuje pojedynczych spraw; staje się recenzentem dla spraw wymagających eskalacji oraz opiekunem reguł biznesowych, które wyznaczają zakres działania agenta. Ekonomika procesu jest zdeterminowana przez jednostkowe koszty decyzji, a nie przez liczbę pracowników.
- **Punkty nadzoru człowieka są osadzone celowo**, nie reaktywnie. W wariancie AI-First udział człowieka w rozstrzyganiu reklamacji wynika z klasyfikacji ryzyka sprawy (np. wartość, klient strategiczny, podejrzenie nadużycia), a nie z domyślnego założenia o nadzorze.
- **Wariant AI-First wymaga znacznie ściślejszego rygoru monitorowania działania systemu**. Każde wywołanie modelu oraz każde wywołanie narzędzia jest rejestrowane w dzienniku, każda decyzja zostawia ślad audytowy, a dryf jakości powinien być wykrywany przede wszystkim przez monitorowanie wskaźników jakości, przy czym reklamacje klientów stanowią dodatkowy, a nie podstawowy mechanizm detekcji.

Etap	Nazwa etapu	Symbol	Rekomendacja interpretacyjna	Uzasadnienie
1	Zgłoszenie problemu i klasyfikacja przez GenAI	R (RÓWNOLEGŁY)	Pozostawić R jako typ dominujący, jeżeli kupujący i agent współtworzą opis problemu. Dodać interakcje wtórne: Z, gdy kupujący zatwierdza roszczenie; E, gdy sprawa wymaga doradcy	R opisuje koprodukcję wyniku, ale nie przesądza o nadzorze nad decyzją końcową
2	Założenie wątku i wygenerowanie pierwszej wiadomości	A (AUTONOMICZNY)	Pozostawić A tylko wtedy, gdy założenie i wysłanie wątku nie wymaga akceptacji człowieka. Jeżeli treść roszczenia ma być zatwierdzana przez kupującego, oznaczyć etap jako Z albo dodać Z jako interakcję wtórną	Automatyczne utworzenie technicznego wątku różni się od decyzji lub oświadczenia o skutku procesowym
3	Wymiana komunikacji z monitoringiem SLA	N (NADZOROWANY)	Pozostawić N jako typ dominujący. Dodać E jako interakcję wtórną dla przekroczeń SLA, braku wartościowej odpowiedzi lub wykrycia anomalii	Monitoring pracy agenta i możliwość interwencji odpowiadają logice *human-on-the-loop*
4	Pomoc – bot lub doradca	E (ESKALACJA)	Doprecyzować, że E oznacza mechanizm przekazania sprawy. Odpowiedź botowa może działać w A/N, a pakiet rekomendacji dla doradcy powinien działać w Z	Etap jest złożony i nie powinien być interpretowany jako „czysta eskalacja”
5	Proaktywna kompensacja	Z (ZATWIERDZENIE)	Pozostawić Z jako typ bazowy. A dopuścić wyłącznie dla scenariuszy niskiego ryzyka, niskiej wartości, odwracalnych i objętych silnymi regułami antyfraudowymi	Skutek finansowy i ryzyko nadużyć uzasadniają zatwierdzenie lub silniejszy nadzór człowieka
6	Zamknięcie wątku, NPS i feedback	F (FEEDBACK)	Pozostawić F dla ankiety NPS i uczenia organizacyjnego. Oddzielić automatyczne zamknięcie wątku od feedbacku: zamknięcie może być A/N/Z zależnie od skutku dla klienta i sprzedającego	Feedback nie jest trybem nadzoru nad decyzją o zamknięciu sprawy

8.2. EKONOMIKA TRZECH WARIANTÓW

Założenia ilustracyjne (nie odnoszą się do żadnej konkretnej organizacji): wolumen 200 tys. reklamacji rocznie, średni czas obsługi w wariancie klasycznym 25 minut, w wariancie AI-Ready 12 minut, w wariancie AI-First 1,5 minuty pracy automatycznej na sprawę, przy czym 30% spraw wymaga dodatkowej recenzji człowieka trwającej średnio 4 minuty. Oczekiwany czas obsługi w wariancie AI-First wynosi zatem **1,5 + 0,3 × 4 = 2,7 minuty na sprawę**, co względem wariantu klasycznego oznacza redukcję czasu przetwarzania o około 89%. Jeżeli liczyć wyłącznie pracę człowieka, oszczędność wynosi 0,3 × 4 = 1,2 minuty na sprawę.

W tym hipotetycznym scenariuszu wariant AI-Ready obniża pracochłonność człowieka o około 50%, lecz zachowuje strukturę kosztów opartą na pracy ludzkiej. Wariant AI-First obniża czas przetwarzania sprawy o około 89% (a pracochłonność człowieka o około 95%), wprowadza jednak istotne pozycje kosztowe nieobecne w pierwszych dwóch wariantach: utrzymanie środowiska agentowego, monitorowanie pracy systemu, koszt wywołań inferencyjnych modelu, koszt utrzymania bazy wiedzy, koszt obsługi incydentów oraz koszt zarządzania ryzykiem modelowym. **Ekonomika powinna być więc liczona całościowo dla horyzontu 2-3 lat**, a nie wyłącznie jako oszczędność pracochłonności.

9. KARTA JEDNOSTKI PRACY POZNAWCZEJ

Wprowadzony w opracowaniu aparat pojęciowy – dane wejściowe, dane wyjściowe, kryteria jakości, polityki eskalacji, role kontrolne – wymaga w praktyce konkretnego artefaktu operacyjnego, który spina go w jedną całość. Rolę tę pełni karta jednostki pracy poznawczej. Karta jest narzędziem dokumentacyjnym, projektowanym dla każdej istotnej czynności umysłowej w procesie AI-Ready lub AI-First. Stanowi punkt odniesienia dla architekta rozwiązania, właściciela procesu, właściciela ryzyka modelowego oraz audytora wewnętrznego.

Pole karty	Opis
Nazwa jednostki pracy poznawczej	Krótkie, jednoznaczne określenie czynności (np. „Klasyfikacja reklamacji według kategorii i ryzyka")
Cel	Co ma zostać rozstrzygnięte, wytworzone lub sklasyfikowane w wyniku tej czynności
Dane wejściowe	Dokumenty, dane systemowe, treść wiadomości klienta, parametry sprawy, dane historyczne
Źródła wiedzy	Regulaminy, procedury, polityki, słowniki, historia decyzji, baza precedensów – wraz z hierarchią: źródła wiążące, referencyjne, pomocnicze, historyczne
Reguły decyzyjne	Forma wykonywalna (DMN, tabela decyzyjna, polityka w kodzie) lub udokumentowane kryteria oceny
Wynik	Decyzja, rekomendacja, klasyfikacja, szkic odpowiedzi, dokument lub akcja systemowa
Kryteria jakości	Trafność, kompletność, zgodność z polityką, terminowość, czytelność uzasadnienia
Tryb nadzoru człowieka	Człowiek w pętli, nadzorujący pętlę, zarządzający pętlą lub poza pętlą – z uzasadnieniem wyboru
Typ interakcji człowiek-AI	Dominujący typ operacyjnej współpracy człowieka z agentem w danej jednostce pracy lub etapie procesu: A (autonomiczny), N (nadzorowany), Z (zatwierdzenie), R (równoległy), E (eskalacja) albo F (feedback). Jeżeli etap zawiera więcej niż jeden wzorzec współpracy, należy wskazać typ dominujący oraz interakcje wtórne
Interakcje wtórne i reguły przełączenia	Kryteria przejścia między typami interakcji, w szczególności z trybu autonomicznego lub nadzorowanego do zatwierdzenia, eskalacji albo feedbacku. Reguły powinny odwoływać się do sygnałów jakości i ryzyka: klasy ryzyka sprawy, kompletności danych, jakości źródeł, anomalii, historii błędów, wartości sprawy oraz skutku decyzji
Próg eskalacji	Sygnały jakości i ryzyka uruchamiające recenzję człowieka (klasa ryzyka, kompletność danych, jakość źródeł, anomalie, historia błędów)
Wymogi audytowe	Zakres i okres rejestrowania danych, uprawnienia dostępu do dziennika oraz zdarzenia generujące alerty
Dane osobowe	Kategorie przetwarzanych danych, podstawa prawna, minimalizacja, retencja, pseudonimizacja, prawa osób
Ryzyka i mechanizmy ograniczania	Halucynacje, błędy wyszukiwania, stronniczość, wyciek danych, błędne akcje agenta – wraz z konkretnymi mechanizmami ograniczania ryzyka

Właściciele	Właściciel procesu, właściciel kontekstu (jakość bazy wiedzy), właściciel ryzyka modelowego (zgodność i jakość modelu)

Karta nie zastępuje mapy procesu w notacji BPMN ani modelu decyzyjnego w notacji DMN – uzupełnia je, koncentrując się na wymiarach krytycznych z perspektywy współpracy człowiek-AI: jakości kontekstu, mechanizmach nadzoru, typach interakcji człowiek-AI, regułach przełączenia między nimi, polityce eskalacji oraz wymogach audytowych. **Brak karty dla istotnej jednostki pracy poznawczej często prowadzi do nieporozumień** między zespołem projek-towym, właścicielem procesu, funkcją zgodności i audytem.

10. MINIMALNY STANDARD PROCESU AI-READY

Poniżej przedstawiono minimalny standard ułatwiający ocenę gotowości procesu do paradygmatu AI-Ready, którego spełnienie powinno być warunkiem deklarowania procesu jako AI-Ready w komunikacji wewnętrznej i zewnętrznej. Standard ten nie jest formalnym certyfikatem – jest listą kontrolną pozwalającą uniknąć nadinterpretacji statusu procesu.

1. **Opisane jednostki pracy poznawczej** – dla każdej istotnej czynności umysłowej w procesie istnieje wypełniona karta jednostki pracy poznawczej.
2. **Wersjonowana baza wiedzy** – źródła wiedzy referencyjnej mają oznaczoną wersję, datę aktualizacji oraz odpowiedzialnego właściciela treści.
3. **Zmapowane źródła danych** – znane są źródła danych wejściowych, ich właściciele, jakość, częstotliwość aktualizacji i podstawa prawna przetwarzania.
4. **Reguły decyzyjne w formie wykonywalnej** – polityki i kryteria akceptacji są dostępne w formie umożliwiającej ich automatyczną walidację (DMN, tabela decyzyjna, polityka w kodzie).
5. **Kryteria jakości wyniku** – każda jednostka pracy poznawczej ma zdefiniowane kryteria akceptacji wyniku oraz sposób ich pomiaru.
6. **Tryby nadzoru człowieka i typy interakcji człowiek-AI** – każda istotna decyzja, rekomendacja lub akcja systemowa ma przypisany tryb nadzoru człowieka proporcjonalny do klasy ryzyka, a każdy etap procesu ma wskazany dominujący typ inter-akcji człowiek-AI: A, N, Z, R, E albo F.
7. **Mechanizm eskalacji i reguły przełączenia interakcji –** eskalacja jest uruchamiana na podstawie wielokryteriowej oceny sygnałów jakości i ryzyka, a nie na podstawie pojedynczego wskaźnika. Dla etapów złożonych należy opisać reguły przełączenia między typami interakcji, w szczególności z trybu autonomicznego lub nadzorowanego do zatwierdzenia, eskalacji albo feedbacku.
8. **Rejestr wywołań i decyzji** – dla wdrożonych komponentów AI każde istotne wywołanie powinno być logowane zgodnie z polityką audytu; dla procesów przygotowywanych do AI-Ready powinien istnieć zaprojektowany mechanizm takiego rejestrowania – z zachowaniem zasad minimalizacji danych, retencji oraz kontroli dostępu do dziennika.
9. **Zestaw testów ewaluacyjnych** – pozwalający na obiektywną ocenę zmian modelu, instrukcji i kontekstu, uruchamiany przed każdym wdrożeniem zmiany.
10. **Zasady ochrony danych i retencji** – spełniające wymogi RODO oraz polityk wewnętrznych, w tym zasadę minimalizacji, ograniczenia celu, pseudonimizacji oraz prawa osób.
11. **Procedura incydentowa** – opisany sposób obsługi incydentów jakościowych i bezpieczeństwa, z dedykowanymi rolami i czasami reakcji.
12. **Zdefiniowani właściciele** – wskazany właściciel procesu, właściciel kontekstu oraz właściciel ryzyka modelowego.

11. KIEDY NIE STOSOWAĆ PARADYGMATU AI-FIRST

Wzbogacenie kryteriów decyzyjnych o jednoznaczne kryteria wykluczające jest niezbędne, by uniknąć przedwczesnych deklaracji architektonicznych. Paradygmat AI-First nie powinien być wybierany, gdy spełniony jest którykolwiek z poniższych warunków:

- decyzje podejmowane w procesie są nieodwracalne lub koszt korekty błędu jest istotnie wyższy niż wartość pojedynczej sprawy, a organizacja nie przewiduje obowiązkowej recenzji człowieka przed wywołaniem skutku decyzji;
- organizacja nie dysponuje wiążącą bazą wiedzy referencyjnej, a jej zbudowanie nie jest częścią planu programu;
- brak jest mechanizmów obiektywnego pomiaru jakości decyzji oraz zestawu testów ewaluacyjnych pozwalających na ich monitorowanie w czasie;
- nie wskazano formalnego właściciela ryzyka modelowego z odpowiednim mandatem zarządczym;
- nie da się zaprojektować skutecznego nadzoru człowieka odpowiedniego do klasy ryzyka systemu w rozumieniu AI Act;
- proces obejmuje decyzje dotyczące osób fizycznych, które mogą wywoływać skutki prawne, a organizacja nie rozstrzygnęła dopuszczalności automatyzacji, podstawy prawnej, ewentualnych wyjątków z art. 22 RODO oraz wymaganych zabezpieczeń dla osoby, której dane dotyczą;
- organizacja nie dysponuje procedurą obsługi incydentów ani sprawdzonym mechanizmem wycofania zmian w warstwie modeli, instrukcji lub kontekstu;
- wynik ekonomiczny programu nie jest dodatni po uwzględnieniu całkowitego kosztu jednostkowego, kosztu zgodności regulacyjnej oraz oczekiwanej straty z błędów.

Wystąpienie któregokolwiek z powyższych warunków nie wyklucza paradygmatu AI-First w przyszłości. **Wskazuje natomiast, że organizacja powinna pozostać na ścieżce adaptacyjnej i konsekwentnie zamykać luki**, zamiast deklarować stan zaawansowania, którego faktycznie nie osiągnęła. Świadomość kryteriów wykluczających jest cechą dojrzałej organizacji wdrożeniowej.

12. RAMY REGULACYJNE I STANDARDY

Wdrożenie procesów AI-Ready i AI-First nie odbywa się w próżni regulacyjnej. Organizacje europejskie funkcjonują w środowisku rosnącej liczby aktów prawnych i standardów branżowych, które kształtują zarówno dopuszczalne sposoby projektowania systemów AI, jak i wymagania dotyczące ich nadzoru, dokumentacji i obsługi incydentów.

Akt / Standard	Zakres i znaczenie dla procesów AI-Ready / AI-First
AI Act (rozporządzenie UE 2024/1689)	Rozporządzenie Parlamentu Europejskiego i Rady ustanawiające zharmonizowane przepisy dotyczące sztucznej inteligencji – pierwszy horyzontalny akt UE w tej dziedzinie. Wprowadza reżim oparty na ryzyku, na który składają się: zakaz określonych praktyk AI, szczególne wymagania dla systemów wysokiego ryzyka (m.in. ocena zdolności kredytowej, zatrudnienie, edukacja, dostęp do usług podstawowych, egzekwowanie prawa, wymiar sprawiedliwości – zgodnie z Załącznikiem III), obowiązki przejrzystości dla wybranych zastosowań (art. 50) oraz obowiązki ogólne – w tym kompetencje AI (*AI literacy*). Dla projektowania procesów AI-Ready i AI-First istotne jest nie tylko ustalenie, czy system jest systemem wysokiego ryzyka, lecz także określenie roli organizacji: dostawcy, podmiotu stosującego, importera, dystrybutora lub producenta produktu. Akt stosowany jest etapowo: zakazy określonych praktyk i obowiązek kompetencji AI obowiązują od 2 lutego 2025 r.; przepisy o ładzie zarządczym oraz obowiązki dostawców modeli ogólnego przeznaczenia (GPAI) – od 2 sierpnia 2025 r. Pierwotne terminy stosowania obowiązków dla systemów wysokiego ryzyka – 2 sierpnia 2026 r. (Załącznik III) i 2 sierpnia 2027 r. (Załącznik I). 7 maja 2026 r. Parlament Europejski i Rada UE osiągnęły wstępne porozumienie polityczne w sprawie pakietu Digital Omnibus, przewidującego m.in. przesunięcie tych terminów odpowiednio do 2 grudnia 2027 r. i 2 sierpnia 2028 r. oraz przesunięcie terminu stosowania obowiązków technicznego oznaczania treści generowanych przez AI (art. 50 ust. 2) do 2 grudnia 2026 r. Szczegółowy zakres tego obowiązku należy odczytywać zgodnie z ostatecznie przyjętym brzmieniem aktu. Zmiany wynikające z porozumienia politycznego wymagają formalnego przyjęcia w procedurze ustawodawczej; zgodnie z komunikatami dotyczącymi porozumienia formalne przyjęcie zmian powinno zostać poddane monitorowaniu w toku dalszej procedury legislacyjnej. Do czasu przyjęcia aktu zmieniającego obowiązują terminy wynikające z aktualnego brzmienia AI Act
ISO/IEC 42001:2023	Międzynarodowy standard systemu zarządzania sztuczną inteligencją (AI Management System). Określa wymagania dotyczące ustanawiania, wdrażania, utrzymywania i ciągłego doskonalenia systemu zarządzania AI w organizacji, obejmując polityki, cele, role, procesy, ocenę ryzyka, monitorowanie i ocenę skuteczności. Może służyć jako systemowa rama zarządzania AI, umożliwiająca uporządkowanie polityk, ról, ryzyk i mechanizmów monitorowania, a także wspierać mapowanie wymogów AI Act i praktyk NIST AI RMF – nie zastępuje jednak klasyfikacji systemu w rozumieniu AI Act, oceny zgodności ani analizy prawnej
NIST AI Risk Management Framework	Dobrowolne ramy zarządzania ryzykiem AI rozwijane przez Narodowy Instytut Standaryzacji i Technologii Stanów Zjednoczonych (NIST). Definiują cztery funkcje: *Govern* (ład zarządczy), *Map* (identyfikacja kontekstu i ryzyk), *Measure* (pomiar) oraz *Manage* (zarządzanie ryzykiem). Promują wielowymiarowe, cykliczne podejście do ryzyka uwzględniające zaufanie w projektowaniu, rozwoju, wdrażaniu i ocenie systemów AI. Stosowane podejście oparte na wielokryteriowej ocenie

	sygnałów jakości i ryzyka jest zgodne z duchem ram NIST, lecz nie stanowi formalnie wymaganej przez nie metody. Często wykorzystywane jako uzupełnienie wymogów ISO/IEC 42001 oraz AI Act, w szczególności w organizacjach o globalnej skali działania
DORA (rozporządzenie UE 2022/2554)	Rozporządzenie o cyfrowej odporności operacyjnej sektora finansowego, obowiązujące od 17 stycznia 2025 r. Nakłada na banki, ubezpieczycieli, firmy inwestycyjne i innych uczestników rynku finansowego wymagania dotyczące zarządzania ryzykiem ICT, testowania odporności, obsługi i zgłaszania incydentów oraz nadzoru nad krytycznymi dostawcami zewnętrznymi usług ICT. W kontekście procesów AI-First i AI-Ready DORA ma zastosowanie wówczas, gdy model, platforma AI, infrastruktura chmurowa lub warstwa orkiestracji stanowią usługę ICT wspierającą proces biznesowy podmiotu finansowego – szczególnie jeśli usługa ta jest klasyfikowana jako krytyczna lub istotna dla działania procesu. Wymogi DORA są komplementarne wobec AI Act i nakładają się przede wszystkim w obszarach zarządzania ryzykiem zewnętrznych dostawców, ciągłości działania i zgłaszania incydentów
RODO (rozporządzenie UE 2016/679)	Ogólne rozporządzenie o ochronie danych osobowych obowiązujące od 25 maja 2018 r. Dla procesów AI-Ready i AI-First istotne są przede wszystkim art. 22 (prawo do niepodlegania decyzji opartej wyłącznie na zautomatyzowanym przetwarzaniu, w tym profilowaniu, wywołującej skutki prawne lub w podobny sposób istotnie wpływającej na osobę), art. 35 (ocena skutków dla ochrony danych (DPIA) wymagana, gdy dany rodzaj przetwarzania może powodować wysokie ryzyko naruszenia praw lub wolności osób fizycznych, w szczególności przy systematycznej i kompleksowej ocenie czynników osobowych opartej na zautomatyzowanym przetwarzaniu), wymóg podstawy prawnej przetwarzania, zasada minimalizacji danych, ograniczenie celu, ograniczenie czasu przechowywania oraz prawa osób (informacja, dostęp, sprostowanie, usunięcie). Dla procesów wykorzystujących AI do podejmowania decyzji dotyczących osób fizycznych RODO może nakładać silniejsze ograniczenia operacyjne niż AI Act, zwłaszcza w zakresie zautomatyzowanego podejmowania decyzji, podstawy prawnej przetwarzania i praw osób, których dane dotyczą
Krajowe wytyczne sektorowe	W Polsce – wytyczne KNF dla sektora finansowego (m.in. dotyczące outsourcingu chmurowego i bezpieczeństwa systemów informatycznych), rekomendacje UODO dotyczące ochrony danych osobowych, wytyczne resortowe dla sektora publicznego. Stanowią uszczegółowienie wymogów horyzontalnych dla sektorów regulowanych

12.1. PRAKTYCZNE IMPLIKACJE DLA PROJEKTOWANIA PROCESÓW

Dla organizacji regulowanych – banków, ubezpieczycieli, dostawców usług zdrowotnych, sektora publicznego – rekomendowane jest **osadzenie procesów AI-Ready i AI-First w systemie zarządzania AI zgodnym z ISO/IEC 42001**, obejmującym polityki, cele, procesy, role, ryzyka, monitorowanie i ciągłe doskonalenie. Standard ten może stanowić ramę systemu zarządzania AI i ułatwiać mapowanie wymogów AI Act oraz praktyk NIST AI RMF, ale nie zastępuje klasyfikacji systemu, oceny zgodności ani analizy prawnej.

Z perspektywy projektowania procesu kluczowe są cztery implikacje regulacyjne:

1. **Klasyfikacja ryzyka systemu AI wynika z przepisów, przeznaczenia systemu i kontekstu użycia.** Decyzją projektową jest natomiast dostosowanie architektury, zakresu automatyzacji i trybu nad-

zoru do tej klasyfikacji. Wybór paradygmatu (AI-Ready lub AI-First) oraz trybu nadzoru człowieka powinien wynikać między innymi z klasyfikacji systemu w rozumieniu AI Act.

2. **Audytowalność musi być wbudowana w architekturę.** W przypadku systemów objętych odpowiednimi wymogami AI Act – w szczególności systemów wysokiego ryzyka – logowanie zdarzeń, dokumentacja techniczna oraz śledzenie zmian modeli i instrukcji powinny być projektowane jako element zgodności regulacyjnej, a nie tylko jako praktyka operacyjna. Logowanie powinno być projektowane z poszanowaniem zasad minimalizacji danych, retencji, pseudonimizacji oraz kontroli dostępu do dziennika.
3. **Adekwatny nadzór człowieka jest wymogiem AI Act.** Dla systemów wysokiego ryzyka niedopuszczalny jest brak realnego, skutecznego i udokumentowanego nadzoru człowieka; zakres udziału człowieka w pojedynczej decyzji powinien wynikać z oceny ryzyka, przeznaczenia systemu, przepisów sektorowych oraz wymagań AI Act i RODO.
4. **Kompetencje AI (*AI literacy*) są obowiązkiem prawnym.** Od 2 lutego 2025 r. dostawcy i podmioty stosujące systemy AI są zobowiązani do podejmowania działań zapewniających odpowiedni poziom kompetencji AI u personelu oraz innych osób zajmujących się obsługą i wykorzystaniem systemów AI w ich imieniu – co wpływa na wymagania szkoleniowe w obu paradygmatach.

12.2. ROLE REGULACYJNE W PROCESACH AI-READY I AI-FIRST

AI Act różnicuje obowiązki w zależności od roli, jaką organizacja pełni względem systemu AI. Dla projektowania procesów AI-Ready i AI-First istotne jest świadome zidentyfikowanie tej roli, ponieważ ta sama organizacja może występować w różnych rolach w odniesieniu do różnych systemów:

- **Dostawca (*provider*)** – podmiot, który opracowuje system AI lub zleca jego opracowanie i wprowadza go do obrotu lub oddaje do użytku pod własną nazwą lub znakiem towarowym; ponosi pełną odpowiedzialność za zgodność systemu z AI Act, w tym za zarządzanie ryzykiem, dokumentację techniczną, ocenę zgodności i znakowanie CE.
- **Podmiot stosujący system AI (*deployer*)** – podmiot, który używa systemu AI pod swoją kontrolą w ramach działalności zawodowej lub organizacyjnej, z wyłączeniem użycia osobistego i nie profesjonalnego; odpowiada między innymi za nadzór człowieka, monitorowanie działania, prowadzenie rejestrów oraz informowanie osób objętych decyzją systemu.
- **Importer i dystrybutor** – podmioty wprowadzające system AI dostawcy spoza Unii Europejskiej do obrotu lub udostępniające go na rynku unijnym; ich obowiązki obejmują weryfikację zgodności, dokumentacji oraz znakowania CE.
- **Producent produktu** – podmiot wprowadzający na rynek produkt zawierający system AI jako element bezpieczeństwa, w przypadku którego AI Act nakłada obowiązki uzupełniające wobec sektorowego prawa harmonizacyjnego.

Z perspektywy procesu AI-First rozróżnienie ról ma znaczenie praktyczne: organizacja może względem zakupionego systemu AI pełnić rolę podmiotu stosującego, lecz **stać się dostawcą** w rozumieniu AI Act, jeżeli wprowadza system do obrotu lub oddaje go do użytku pod własną nazwą lub znakiem towarowym bądź dokonuje istotnej modyfikacji systemu w sposób wpływający na jego przeznaczenie lub zgodność regulacyjną. Ta zmiana statusu pociąga za sobą istotnie szerszy zakres obowiązków regulacyjnych i powinna być ujęta w decyzji architektonicznej dotyczącej rekonstrukcji procesu.

13. REKOMENDACJE DLA LIDERÓW TRANSFORMACJI

Poniższe zalecenia mają charakter syntetyczny i ekspercki. Opierają się na analizie publicznie dostępnych doświadczeń wdrożeniowych w polskich i międzynarodowych organizacjach z sektora finansowego, ubezpieczeniowego, energetycznego, farmaceutycznego oraz administracji publicznej. Są skierowane do liderów odpowiedzialnych za decyzje strategiczne dotyczące projektowania procesów następnej generacji.

13.1. REKOMENDACJE STRATEGICZNE

1. **Nie należy deklarować organizacji jako „AI-First" przed uruchomieniem co najmniej jednego procesu AI-Ready w środowisku produkcyjnym.** Deklaracje strategiczne wyprzedzające zdolności wykonawcze generują ryzyko utraty wiarygodności programu transformacji wewnątrz organizacji, wobec klientów i wobec regulatora.
2. **Inwestycje w architekturę informacyjną powinny poprzedzać inwestycje w warstwę modeli.** Jakość bazy wiedzy, dokumentacji procesowej i danych historycznych determinuje pułap możliwy do osiągnięcia przez najlepszy nawet model.
3. **Procesy do paradygmatu AI-First należy wybierać selektywnie.** W horyzoncie pierwszych 18 miesięcy rekomenduje się wybór dwóch do trzech procesów flagowych. Pozwala to zbudować kompetencje zespołów i wzorce architektoniczne, które następnie skalują się na kolejne procesy. Dokładna liczba tych procesów powinna być pochodną dostępnego budżetu, kompetencji wewnętrznych, profilu ryzyka oraz wolumenu spraw w procesach kandydackich.
4. **Ład zarządczy należy traktować jako element architektury, a nie jako warstwę nadbudowaną.** Polityki użycia danych, rejestr modeli i procedury incydentowe powinny być projektowane równolegle z procesem, a nie dodawane po jego uruchomieniu.
5. **Procesy należy osadzać w systemie zarządzania AI zgodnym z ISO/IEC 42001.** Dla organizacji regulowanych jest to nie tylko najlepsza praktyka, lecz także ścieżka uwiarygodnienia wobec audytora i regulatora.

3.2. REKOMENDACJE OPERACYJNE

1. **Dla każdego procesu wspieranego przez AI należy zbudować zestaw testów ewaluacyjnych.** Bez obiektywnej miary jakości każda decyzja o aktualizacji modelu lub instrukcji opierać się będzie na niezweryfikowanym założeniu jakościowym.
2. **Należy wprowadzić kontrolowane wdrożenia kanarkowe oraz procedury wycofywania zmian.** Każda zmiana modelu, instrukcji lub konfiguracji bazy wiedzy powinna przechodzić przez kontrolowane uruchomienie zmiany na ograniczonym, reprezentatywnym fragmencie ruchu produkcyjnego (wdrożenie kanarkowe) z możliwością szybkiego wycofania.
3. **Należy zdefiniować role właściciela kontekstu oraz właściciela ryzyka modelowego.** Klasyczna rola właściciela procesu nie obejmuje pełnej odpowiedzialności za jakość warstwy informacyjnej i polityki użycia AI.
4. **Należy wprowadzić cykliczny przegląd procesów AI-Ready pod kątem gotowości do AI-First.** Nie każdy proces dojrzeje do AI-First, ale brak systematycznego przeglądu skutkuje przegapieniem okazji do transformacji.

5. **Mechanizmy bezpieczeństwa agentowego należy wbudowywać od pierwszej iteracji.** Wstrzyknięcie instrukcji, zatrucie bazy wiedzy oraz wyprowadzenie danych przez narzędzia stanowią podstawowe wektory zagrożeń, które należy uwzględniać od pierwszej iteracji projektu, a nie traktować jako zagadnienia eksperckie do rozważenia w późniejszej fazie wdrożenia.

13.3. REKOMENDACJE KOMPETENCYJNE

W obu paradygmatach krytycznym czynnikiem powodzenia są kompetencje zespołów projektowych. Wymagana jest zmiana w trzech wymiarach: poszerzenie zakresu kompetencji analityka biznesowego o pracę z modelami i danymi nieustrukturyzowanymi; wprowadzenie nowej roli architekta rozwiązań AI łączącej spojrzenie procesowe, danowe i modelowe; spełnienie wymogu kompetencji AI (*AI literacy*) zgodnie z art. 4 AI Act – obejmującego personel oraz osoby działające na rzecz organizacji, a zaangażowane w obsługę i wykorzystanie systemów AI.

Rola architekta rozwiązań AI nie jest tożsama z rolą inżyniera uczenia maszynowego – koncentruje się na projektowaniu współpracy człowiek-AI w ramach procesu, na klasyfikacji ryzyka, polityce eskalacji, mechanizmach kontroli i wyborze trybu nadzoru, a nie na trenowaniu modeli.

14. PODSUMOWANIE

AI-Ready oraz AI-First nie są dwoma punktami na skali dojrzałości – są dwoma odrębnymi paradygmatami projektowania procesów, między którymi istnieje fundamentalna różnica architektoniczna. Pierwszy z nich polega na adaptacji procesu do bezpiecznej, mierzalnej i kontrolowanej współpracy z AI; drugi – na zaprojektowaniu procesu według zasady, że komponent AI jest domyślnym wyborem do wykonania pracy poznawczej, o ile nie wykluczają tego ryzyko, regulacje, koszt błędu lub potrzeba eksperckiego osądu.

Wybór między paradygmatami nie jest rozstrzygnięciem uniwersalnym. Determinują go cechy procesu (wolumen, koszt jednostkowy, profil ryzyka), dojrzałość organizacji (stan danych, kompetencje, ład zarządczy), wymagania regulacyjne oraz horyzont strategiczny. Dla większości procesów w polskich organizacjach właściwą ścieżką pozostaje sekwencyjne dojrzewanie: As-Is → AI-Augmented → AI-Ready. Paradygmat AI-First jest właściwym wyborem dla wybranych procesów flagowych, w których oczekiwana wartość netto przedsięwzięcia jest dodatnia, a profil ryzyka pozwala na spełnienie wymogów regulacyjnych.

Najpoważniejsze ryzyko zidentyfikowane w praktyce wdrożeniowej polega nie na nadmiernej ostrożności, lecz na przedwczesnym deklarowaniu transformacji do modelu AI-First przy jednoczesnym braku ustabilizowanych procesów AI-Ready w środowisku produkcyjnym. Powstaje wówczas rozdźwięk między ambicją strategiczną a zdolnością wykonawczą, który podważa wiarygodność programu transformacji - zarówno wewnątrz organizacji, jak i wobec klientów, audytora oraz regulatora.

Najtrwalszą wartością paradygmatu **AI-Ready** jest budowa **dyscypliny procesowej, jakości danych i dojrzałości ładu zarządczego** – fundamentów, bez których proces AI-First pozostanie prototypem. Najważniejszą wartością paradygmatu **AI-First** jest natomiast możliwość **trwałego obniżenia kosztu jednostkowego powtarzalnej pracy poznawczej oraz zaprojektowania nowych modeli operacyjnych**. Rolą lidera transformacji jest zaprojektowanie portfela procesów, w którym oba paradygmaty współistnieją i wzmacniają się wzajemnie w ramach systemu zarządzania AI dostosowanego do wymogów AI Act, ISO/IEC 42001 i ram NIST AI RMF.

SŁOWNIK KLUCZOWYCH POJĘĆ

Pojęcie	Definicja
Agent	Komponent zdolny do autonomicznego planowania i realizacji wielokrokowych zadań w środowisku z dostępem do narzędzi. Ma określony stan roboczy, podejmuje decyzje o sekwencji działań, iteruje w odpowiedzi na wyniki pośrednie
RAG agentowy (*Agentic RAG*)	Architektura wyszukiwania, w której agent autonomicznie decyduje o sposobie pobierania kontekstu, łącząc wyszukiwanie semantyczne, wywoływanie narzędzi i samodzielne formułowanie zapytań pomocniczych
AI Act	Rozporządzenie UE 2024/1689 – pierwszy w Unii Europejskiej horyzontalny akt regulujący sztuczną inteligencję, klasyfikujący systemy według poziomu ryzyka. Stosowany stopniowo od 2 lutego 2025 roku
Kompetencje AI (*AI literacy*)	Kompetencje pozwalające na zrozumienie działania systemów AI, krytyczną ocenę ich wyników oraz bezpieczne korzystanie z nich. Obowiązek wprowadzony przez AI Act z mocą obowiązującą od 2 lutego 2025 roku
Brama decyzyjna	Punkt w ścieżce transformacji, w którym dalsze przejście wymaga spełnienia określonych warunków technicznych, organizacyjnych i kompetencyjnych
DORA	Rozporządzenie UE 2022/2554 o cyfrowej odporności operacyjnej sektora finansowego stosowane od 17 stycznia 2025 roku. Komplementarne wobec AI Act w zakresie ryzyk technologicznych i zarządzania dostawcami
GraphRAG	Wariant architektury RAG, w którym warstwa wyszukiwania wykorzystuje grafową reprezentację wiedzy, umożliwiając zapytania o relacje między bytami
Oparcie odpowiedzi na źródłach (*grounding*)	Stopień, w jakim odpowiedź modelu jest oparta na wiarygodnych źródłach, możliwych do wskazania i zweryfikowania
Człowiek w pętli decyzyjnej (*human-in-the-loop*)	Tryb nadzoru, w którym człowiek zatwierdza wynik przed wywołaniem skutku decyzji
Człowiek nadzorujący pętlę (*human-on-the-loop*)	Tryb nadzoru, w którym człowiek monitoruje pracę systemu i ma uprawnienia interwencji lub zatrzymania
Człowiek zarządzający pętlą decyzyjną (*human-over-the-loop*)	Tryb nadzoru, w którym człowiek zarządza politykami, testami, wyjątkami i audytem, nie nadzorując pojedynczych decyzji
Bez udziału człowieka w pojedynczej decyzji (*human-out-of-the-loop*)	Tryb bez udziału człowieka w pojedynczej decyzji operacyjnej (przy zachowaniu nadzoru projektowego, monitoringu, audytu i procedur incydentowych), dopuszczalny jedynie dla decyzji niskiego ryzyka po formalnej ocenie
Typ interakcji człowiek-AI	Operacyjny sposób współpracy człowieka z agentem AI w konkretnym etapie procesu. Typ interakcji opisuje organizację pracy, nie zastępuje trybu nadzoru nad decyzją
Dominujący typ interakcji	Główny symbol A/N/Z/R/E/F przypisany do etapu procesu
Interakcje wtórne	Dodatkowe wzorce współpracy człowiek-AI występujące w etapie procesu obok typu dominującego. Przykładowo etap oznaczony jako E (Eskalacja) może zawierać także przygotowanie rekomendacji do zatwierdzenia w trybie Z

Eskalacja jako mechanizm przełączenia	Reguła przekazania sprawy do człowieka po spełnieniu kryteriów ryzyka, jakości, niejednoznaczności, braku danych, anomalii lub przekroczenia progu decyzyjnego. Eskalacja nie jest samodzielnym trybem nadzoru, lecz prowadzi zwykle do *human-in-the-loop* albo *human-on-the-loop*
Feedback jako mechanizm kalibracji	Informacja zwrotna od człowieka dotycząca jakości wyniku agenta, wykorzystywana do korekty reguł, instrukcji, źródeł wiedzy, testów ewaluacyjnych, metryk jakości lub konfiguracji systemu. Feedback nie jest nadzorem nad pojedynczą decyzją, ale może wspierać *human-over-the-loop*
ISO/IEC 42001	Międzynarodowy standard systemu zarządzania sztuczną inteligencją (*AI Management System*), opublikowany w 2023 roku
***Vendor Lock-in* (uzależnienie od dostawcy)**	Uzależnienie od konkretnego dostawcy modelu, platformy, formatów danych lub mechanizmów orkiestracji, którego zmiana wiąże się z istotnym kosztem migracji, przebudowy architektury lub renegocjacji warunków handlowych
Karta jednostki pracy poznawczej	Artefakt operacyjny opisujący istotną czynność umysłową w procesie: dane wejściowe, źródła wiedzy, reguły decyzyjne, wynik, kryteria jakości, tryb nadzoru, próg eskalacji, wymogi audytowe, dane osobowe, ryzyka oraz właścicieli.
Model Context Protocol (MCP)	Otwarty protokół integracji modeli AI z narzędziami i źródłami danych, ogłoszony przez firmę Anthropic w listopadzie 2024 r, a 9 grudnia 2025 r. przekazany do Agentic AI Foundation – funduszu celowego pod auspicjami Linux Foundation, którego współzałożycielami są Anthropic, Block oraz OpenAI, a wspierającymi członkami między innymi Google, Microsoft, AWS, Cloudflare i Bloomberg. Ogranicza uzależnienie od dostawcy na poziomie integracyjnym, zwiększa znaczenie bezpieczeństwa narzędzi i serwerów udostępniających usługi w protokole, w tym ryzyk wstrzyknięcia instrukcji pośredniego, podstawienia narzędzi i wycieku tokenów
NIST AI RMF	AI Risk Management Framework – amerykańskie ramy zarządzania ryzykiem AI, oparte na funkcjach *Govern, Map, Measure, Manage*
Wskaźnik korekt rekomendacji (*override rate*)	Wskaźnik korekt – odsetek rekomendacji modelu, które zostały zmienione lub odrzucone przez recenzenta. Kluczowa metryka jakości procesu AI-Ready i AI-First.
Próg eskalacji oparty na wielokryteriowej ocenie sygnałów	Wartość progowa wyliczana z kilku sygnałów (klasa ryzyka, kompletność danych, jakość źródeł, zgodność z regułami, anomalie, historia błędów), wykorzystywana do decyzji o przekazaniu sprawy do recenzji człowieka
RAG (*Retrieval-Augmented Generation*)	Architektura, w której generatywny model językowy odpowiada na podstawie kontekstu pobranego z wiarygodnej bazy wiedzy, a nie wyłącznie z parametrów modelu
Rejestr modeli i systemów AI	Element ładu zarządczego gromadzący informacje o wszystkich modelach i systemach AI używanych w organizacji: ich wersjach, zastosowaniach, właścicielach, profilu ryzyka i wynikach ewaluacji
RODO (GDPR)	Rozporządzenie UE 2016/679 o ochronie danych osobowych, obowiązujące od 25 maja 2018 r. Dla procesów AI istotne są: art. 22 (decyzje zautomatyzowane), art. 35 (DPIA), zasady minimalizacji i ograniczenia celu oraz prawa osób, których dane dotyczą
Stronniczość algorytmiczna	Systematyczna tendencja modelu lub procesu decyzyjnego do faworyzowania lub gorszego traktowania określonych grup, kategorii spraw lub kanałów kontaktu, wynikająca z danych treningowych, projektu modelu lub mechanizmów eskalacji

System wieloagentowy	Architektura, w której wiele wyspecjalizowanych agentów AI współpracuje nad realizacją zadania - każdy z odrębną rolą, narzędziami i kontekstem
Wstrzyknięcie instrukcji (*prompt injection*)	Klasa ataków, w których dane wejściowe modelu zawierają instrukcje przejmujące kontrolę nad jego zachowaniem. Forma bezpośrednia (w danych użytkownika) i pośrednia (w pobieranych dokumentach)
Wywoływanie narzędzi (*tool use*)	Mechanizm, w którym aplikacja, orkiestrator lub środowisko agentowe umożliwia modelowi inicjowanie wywołań zewnętrznych narzędzi – takich jak API, funkcje lub bazy danych – zgodnie z nadanymi uprawnieniami i politykami bezpieczeństwa
Zestaw testów ewaluacyjnych	Zbiór testów regresyjnych pozwalający na obiektywną ocenę jakości pracy modelu lub agenta po zmianach wersji modelu, kontekstu lub instrukcji
Właściciel kontekstu	Rola odpowiedzialna za jakość, aktualność i kompletność warstwy informacyjnej dostarczanej modelowi w ramach procesu
Właściciel ryzyka modelowego	Rola odpowiedzialna za zgodność modelu z wymogami regulacyjnymi i jakościowymi, monitorowanie ryzyk modelowych, decyzje o wdrożeniu zmian oraz reakcję na incydenty jakościowe

BIBLIOGRAFIA (WYBRANE POZYCJE)

Lp.	Autorzy	Tytuł	Miejsce i data publikacji	Krótki opis	Link do artykułu
1	Ali, Wajid; Khan, Abdul Zahid	Factors influencing readiness for artificial intelligence: a systematic literature review	Data Science and Management, vol. 8, nr 2, Elsevier/KeAi, 2025	Artykuł syntetyzuje czynniki gotowości organizacyjnej do wdrażania sztucznej inteligencji. Publikacja stanowi punkt odniesienia dla warstwy AI-Ready na poziomie organizacji: infrastruktury IT, kompetencji, wsparcia kierownictwa, kultury, jakości danych i zasobów	Otwórz www.sciencedirect.com/science/article/pii/S2666764924000511
2	Calvanese, Diego; Casciani, Angelo; De Giacomo, Giuseppe; Dumas, Marlon; Fournier, Fabiana; Kampik, Timotheus; La Malfa, Emanuele; Limonad, Lior; Marrella, Andrea; Metzger, Andreas; Montali, Marco; Amyot, Daniel; Fettke, Peter; Polyvyanyy, Artem; Rinderle-Ma, Stefanie; Sardina, Sebastian; Tax, Niek; Weber, Barbara	Agentic Business Process Management: A Research Manifesto	Information Systems, Elsevier, 2026; wersja preprint: arXiv, marzec/kwiecień 2026	Manifest badawczy definiuje agentowe zarządzanie procesami biznesowymi jako rozszerzenie BPM dla organizacji, w których autonomiczne agenty wykonują działania procesowe w kontrolowanych ramach	Otwórz arxiv.org/abs/2603.18916
3	Calvanese, Diego; De Giacomo, Giuseppe; Kampik, Timotheus; Lesperance, Yves; Marrella, Andrea; Matta, Andrea	Autonomy in Business Process Execution: Why We Need First-Class Abstractions for Goals and Normative Frames	Proceedings of the 4th International Workshop on Process Management in the AI Era (PMAI 2025), CEUR Workshop Proceedings, vol. 4087, 2025	Artykuł wskazuje potrzebę traktowania celów oraz ram normatywnych jako obiektów pierwszej klasy w projektowaniu autonomicznych procesów. Publikacja wzmacnia fundament pojęciowy dla projektowania ograniczeń, polityk i barier ochronnych (*guardrails*) w procesach AI-First	Otwórz ceur-ws.org/Vol-4087/paper6.pdf
4	De Giacomo, Giuseppe; Kampik, Timotheus; Kirchdorfer, Lukas; Montali, Marco; Weinhuber, Christoph	Formal Foundations of Agentic Business Process Management	arXiv, kwiecień 2026	Praca rozwija formalne podstawy agentowego BPM. Wskazuje, że wykonanie procesu może być napędzane przez wielu autonomicznych decydentów, a specyfikacja procesu musi obejmować jawne cele, założenia i ograniczenia autonomii agentów	Otwórz arxiv.org/abs/2604.17347

5	Dumas, Marlon; Milani, Fredrik; Chapela-Campa, David	Agentic Business Process Management Systems	Business Process Management Workshops 2025, Springer, 2026; wersja preprint: arXiv, styczeń 2026	Artykuł przedstawia architektoniczną wizję systemów A-BPMS, które łączą autonomię, rozumowanie i uczenie się w zarządzaniu oraz wykonywaniu procesów. Publikacja przedstawia rozróżnienie między klasyczną automatyzacją a autonomią procesową	Otwórz arxiv.org/abs/2601.18833
6	Elyasaf, Achiya; Metzger, Andreas; Sardina, Sebastian; Senderovich, Arik; Serral Asensio, Estefanía; Tax, Niek	Toward Self-Modifying Autonomous Business Process Systems	Proceedings of the 4th International Workshop on Process Management in the AI Era (PMAI 2025), CEUR Workshop Proceedings, vol. 4087, 2025	Artykuł analizuje autonomiczne systemy procesowe zdolne do modyfikowania własnego zachowania w czasie. Publikacja jest istotna dla zaawansowanego wariantu AI-First, w którym proces nie tylko wykonuje zadania autonomicznie, ale także adaptuje się do zmian otoczenia	Otwórz ceur-ws.org/Vol-4087/paper1-Long.pdf
7	Fettke, Peter; Fournier, Fabiana; Limonad, Lior; Metzger, Andreas; Rinderle-Ma, Stefanie; Weber, Barbara	XABPs: Towards eXplainable Autonomous Business Processes	Proceedings of the 4th International Workshop on Process Management in the AI Era (PMAI 2025), CEUR Workshop Proceedings, vol. 4087, 2025	Publikacja rozwija pojęcie wyjaśnialnych autonomicznych procesów biznesowych. Tekst jest szczególnie ważny dla ładu, audytowalności, odpowiedzialności, zgodności regulacyjnej i zarządzania zaufaniem w procesach realizowanych przez AI	Otwórz ceur-ws.org/Vol-4087/paper2-Long.pdf
8	Fournier, Fabiana; Limonad, Lior; David, Yuval	Agentic AI Process Observability: Discovering Behavioral Variability	Proceedings of the 4th International Workshop on Process Management in the AI Era (PMAI 2025), CEUR Workshop Proceedings, vol. 4087, 2025; preprint: arXiv, maj 2025	Artykuł dotyczy obserwowalności procesów realizowanych przez agentów AI oraz wykrywania zmienności ich zachowania. Publikacja omawia praktyczne projektowanie monitoringu, debugowania, analizy trajektorii i kontroli niedeterministycznych zachowań agentów	Otwórz ceur-ws.org/Vol-4087/paper3-Long.pdf
9	Montali, Marco; Comuzzi, Marco; Teinemaa, Irene; Amyot, Daniel; Dumas, Marlon	Towards Conversational Actionability in AI-Augmented Business Process Management Systems	Proceedings of the 4th International Workshop on Process Management in the AI Era (PMAI 2025), CEUR Workshop Proceedings, vol. 4087, 2025	Artykuł opisuje systemy AI-augmented BPMS jako procesowo świadome systemy informacyjne, które dynamicznie dostosowują przepływy wykonania na podstawie stanu, kontekstu i celów biznesowych. Publikacja jest istotna dla przejścia od asystowania do działania procesowego	Otwórz ceur-ws.org/Vol-4087/paper8.pdf
10	Schmidt, Rainer; Alt, Rainer; Zimmermann, Alfred	Agentic AI Readiness: A Process-Oriented Assessment Framework	Proceedings of the 59th Hawaii International Conference on System	Artykuł przedstawia procesowo zorientowaną ramę oceny gotowości do wykorzystania agentów AI. Publikacja jest przydatna do	Otwórz scholarspace.manoa.hawaii.

	Sciences (HICSS 2026), Lahaina, s. 4021-4030, styczeń 2026	budowy modelu oceny gotowości procesów do wdrożenia agentów AI i rozwiązań AI-First	edu/bitstreams/8610b05c-71c4-49e1-a8eb-c4a9d9abb968/download