Playbook „AI Tool Selection” dla przedsiębiorstwa

Executive Summary

Ten dokument definiuje i opisuje gotowy do wdrożenia **Playbook „AI Tool Selection”** – wspólny sposób wyboru technologii AI (GenAI, RAG, agenci, klasyczny ML, CV/OCR, optymalizacja, reguły/automatyzacja) dla procesów biznesowych, IT i operacyjnych.

Playbook opiera się na dobrych praktykach projektowania playbooków operacyjnych w inżynierii oprogramowania, cyberbezpieczeństwie i zarządzaniu ryzykiem AI, zaczerpniętych m.in. z praktyk SOC/incident response (rozróżnienie playbook vs runbook), materiałów o „principle playbooks”, standardów NIST AI RMF oraz ISO/IEC 42001 (AI Management System).

Playbook jest zaprojektowany tak, aby:

- Był **operacyjny**: zawiera jasno opisany workflow decyzji (intake → triage → wybór podejścia → MVI → risk & controls → implementacja → release → monitoring → retrospektywa).
- Zapewniał **spójność decyzji** między zespołami Biznes / Product / Data/AI / IT / Security / Legal / Ops, poprzez RACI, governance gates i szablony decyzji.
- Rozróżniał **Fast Path** (niskie ryzyko, szybkie wdrożenie) oraz **Safe Path** (wysokie ryzyko, pełne kontrole i approvals).
- Był kompatybilny z istniejącym **AI Tool Selectorem** (flowchart + macierz + checklisty), przekształcając go w powtarzalny proces decyzyjny.

W drugiej części dokumentu znajdują się kompletne szablony (Markdown) do bezpośredniego użycia w wiki (Confluence/Notion/SharePoint), m.in.: Intake Form, Tool Selection Worksheet, Risk Assessment Checklist, Evaluation Plan, Go/No-Go Checklist, Post-Implementation Review i Playbook Update Request.

Część A – Research: Jak tworzyć playbooki (best practices)

Definicje: playbook vs runbook vs SOP vs policy

Z analizy materiałów o playbookach w cyberbezpieczeństwie, SRE i zarządzaniu incydentami wynika spójny wzorzec rozróżnienia:

- **Playbook**

    - Ramowy, **strategiczny** opis „jak gramy”: jaka jest sekwencja decyzji, kto jest za co odpowiedzialny, jakie są warianty ścieżek i wyjątki.
    - Ma formę **drzewa decyzyjnego / scenariuszy**, obejmuje wiele funkcji (np. technologia, biznes, PR, compliance).
    - Typowy dla cross-funkcyjnych działań (np. incident response, cross-functional agile team playbooks).

- **Runbook**

    - **Szczegółowa instrukcja krok-po-kroku** wykonywana przez konkretny zespół (np. „jak zrestartować usługę X”, „jak przywrócić backup bazy Y”).
    - Ma format checklisty/procedury operacyjnej wykonywanej w powtarzalny sposób.

- **SOP (Standard Operating Procedure)**

    - Sformalizowany dokument opisujący **standardowy sposób wykonania zadania**, często obowiązujący regulacyjnie.
    - SOP bywa nadrzędny do runbooków (SOP mówi „co” i „dlaczego”, runbook „jak dokładnie”).

- **Policy / guideline / framework**

    - **Policy** – wiążące zasady („must/shall”), np. „dane osobowe muszą być szyfrowane w spoczynku”.
    - **Guideline** – zalecenia i dobre praktyki („should”), np. rekomendowane wzorce architektoniczne.
    - **Framework** – struktura logiczna (np. NIST AI RMF z funkcjami Govern/Map/Measure/Manage).

Wniosek dla AI Tool Selection:

- **Playbook** opisuje *kto, kiedy i jak* korzysta z diagramu AI Tool Selector, jakie są ścieżki (Fast/Safe), bramki jakości i ryzyka.
- **Runbooki** i SOP-y opisują szczegółowo, jak np. wykonać RAG eval, jak skonfigurować monitoring, jak zrobić shadow deployment agenta.

Elementy obowiązkowe dobrego playbooka

Z praktyk SOC i incident response, a także z przewodników tworzenia playbooków produktowo-technicznych, wynika kilka powtarzających się elementów:

1. **Cel i zakres** – czemu playbook służy, dla jakich sytuacji, a dla jakich nie.

2. **Definicje i słowniczek** – eliminacja niejednoznaczności (np. „co to znaczy high-risk AI", „co to jest agent").
3. **Role i odpowiedzialności (RACI)** – kto jest Accountable, kto Responsible, kto Consulted/ Informed.
4. **Wejścia/wyjścia procesu** – jakie artefakty są wejściem (np. intake form), jakie wyjściem (decision record, blueprint, eval results).
5. **Workflow decyzji z wyjątkami** – główna ścieżka + wyjątki / escalations.
6. **Kryteria jakości i akceptacji** – punkty Go/No-Go, progi metryk.
7. **Kontrola zmian / wersjonowanie** – kto zatwierdza zmiany, jak się je komunikuje.
8. **Metryki skuteczności i adoption** – jak mierzymy, czy playbook działa (czas decyzji, liczba odchyleń, NPS użytkowników playbooka).
9. **Narzędzia/artefakty wspierające** – szablony, checklisty, decision trees, przykłady.

Jak projektować playbooki, żeby były używane

Badania i praktyczne przewodniki nt. „principle playbooks" i playbooków developerskich pokazują, dlaczego wiele playbooków staje się „shelfware" – są zbyt abstrakcyjne, rozproszone i niedopasowane do codziennej pracy.

Kluczowe wnioski:

- **„Fast answers to real questions"** – playbook musi odpowiadać na pytania typu „czy ten typ danych jest wrażliwy?", „czy ten use-case wymaga RAG czy wystarczy prompt?", a nie na poziomie ogólnych haseł.
- **Lokalizacja tam, gdzie pracują zespoły** – publikacja w hubach developerskich (Backstage), Confluence, integracja z ticketingiem (Jira Service Management decision trees).
- **Interaktywne decision trees / forms** – zamiast PDF-ów; formularze, które prowadzą zespół przez kolejne decyzje.
- **Przykłady DO/DON'T** – konkretne, lokalne przykłady z organizacji.
- **Lekkie szkolenia + certyfikacja zrozumienia** – krótkie moduły e-learning, quizy potwierdzające znajomość zasad.
- **Pomiar adoption i friction** – telemetryka korzystania z playbooka, feedback loops, retrospektywy i iteracje.

Specyfika playbooków dla AI / danych

Playbooki dotyczące AI i danych muszą adresować dodatkowe ryzyka, które są centralne w NIST AI RMF, ISO/IEC 42001 oraz EU AI Act:

- **Halucynacje i grounding** – LLM/RAG mogą generować nieprawdziwe odpowiedzi; konieczne są testy halucynacji, wymóg cytowania źródeł i mechanizmy HITL.
- **Prywatność i dane wrażliwe** – klasyfikacja danych, zakaz używania narzędzi bez kontroli nad danymi (shadow AI), privacy-by-design.

- **Bezpieczeństwo i IP** – firma musi kontrolować, jakie dane wychodzą do zewnętrznych modeli, jak chronione jest IP i know-how.
- **Audytowalność i monitoring** – logi promptów, odpowiedzi, użytych narzędzi; inventory systemów AI; monitorowanie driftu danych/modeli.
- **HITL i guardrails** – szczególnie dla high-risk AI (finanse, HR, krytyczne decyzje) niezbędne są mechanizmy nadzoru człowieka oraz polityki guardrails (policy-as-code, limity akcji, sandboxy).
- **Ocena jakości offline i online** – zestawy testów offline (eval sets) + ciągłe monitorowanie jakości i ryzyk w produkcji zgodnie z funkcjami Map/Measure/Manage NIST AI RMF.

Standardy i modele dojrzałości jako wzmocnienie playbooka

- **NIST AI RMF (Govern / Map / Measure / Manage)** – daje strukturę funkcji i kategorii dla zarządzania ryzykiem AI, którą można odwzorować na etapy playbooka (Govern → governance & RACI; Map → intake & triage; Measure → eval; Manage → deployment, monitoring, improvements).
- **ISO/IEC 42001:2023 (AI Management System)** – nakłada wymagania dot. kontekstu organizacji, leadership, risk assessment, operacyjnych kontroli i PDCA dla AI, co przekłada się na potrzebę formalnych playbooków, inventory systemów AI, przeglądów zarządczych i ciągłego doskonalenia.
- **EU AI Act** – wymusza klasyfikację systemów AI wg poziomu ryzyka; high-risk use-cases (np. scoring kredytowy, HR, krytyczna infrastruktura) wymagają dokumentacji, human oversight i zarządzania danymi, co powinno być odzwierciedlone w „Safe Path” playbooka.

Design principles dla Playbooka „AI Tool Selection”

Na bazie powyższych praktyk proponowane zasady projektowe:

1. **Problem-first, nie tech-first** – playbook wymusza start od opisu procesu i „job to be done”, a dopiero potem prowadzi do wyboru typu rozwiązania AI.
2. **Jedna źródłowa wersja decyzji** – każda decyzja technologiczna musi mieć Decision Record (1 strona: kontekst, alternatywy, wybór, uzasadnienie, ryzyka).
3. **Fast Path vs Safe Path** – ścieżka szybkiego działania dla niskiego ryzyka oraz ścieżka pełnej kontroli i governance dla wysokiego ryzyka.
4. **Cross-funkcyjny RACI** – jasno wskazany Business Owner, Product Owner, Data/AI Lead, IT/Architecture, Security, Legal/Compliance, Ops/Run, QA.
5. **Wyraźne „kiedy używać”** – trigger conditions opisane prostymi zasadami (np. „każdy nowy use-case z LLM lub modelem trenowanym na danych klientów przechodzi przez playbook”).

6. **Minimalny narzut** – formularze i checklisty projektowane jako 1–2 strony, które da się wypełnić w 30–60 minut.
7. **Oparcie na istniejącym AI Tool Selectorze** – playbook nie duplikuje logiki technicznej; zamiast tego opisuje, jak i kiedy korzystać z diagramu oraz tabeli.
8. **Guardrails dla AI** – wbudowane zasady dot. danych, prywatności, shadow AI, audytowalności, HITL.
9. **Feedback & iteration** – proces retrospektywy i kanał zgłaszania zmian do playbooka (Playbook Update Request).
10. **Spójność z NIST/ISO/EU AI Act** – mapowanie etapów playbooka na Govern/Map/Measure/Manage i PDCA, aby ułatwić audyt i certyfikację.

Najczęstsze błędy w playbookach i jak ich uniknąć

1. **Zbyt wysoki poziom ogólności**

    - Problem: hasła w stylu „chroń dane klienta” bez definicji danych wrażliwych, progów, przykładów.
    - Jak uniknąć: dodaj sekcję „definicje + przykłady DO/DON’T” dla danych i decyzji.

2. **Brak jednoznacznego właściciela**

    - Problem: nikt nie czuje się odpowiedzialny za utrzymanie i aktualizację playbooka.
    - Jak uniknąć: wyznacz „Playbook Ownera” (np. AI Governance Lead) z mandatem i RACI.

3. **Brak powiązania z realnymi workflowami**

    - Problem: playbook jest w PDF-ie, a zespoły pracują w Jira/ServiceNow/Backstage.
    - Jak uniknąć: osadź playbook jako decision tree w narzędziach, których zespoły używają na co dzień.

4. **Brak mechanizmów feedbacku i iteracji**

    - Problem: playbook się starzeje, a zespoły tworzą własne „plemienne” praktyki.
    - Jak uniknąć: zaplanuj regularne retrospektywy, wskaźniki adoption, formularz Playbook Update Request.

5. **Brak rozróżnienia Fast Path vs Safe Path**

- Problem: te same ciężkie procesy stosuje się do prostego generatora treści i do systemu scoringu kredytowego – prowadzi to albo do omijania playbooka, albo do ryzyka.
- Jak uniknąć: wyraźne progi ryzyka (np. typ danych, stawka błędu, wpływ na klienta) kierujące do odpowiedniej ścieżki.

6. **Brak mierników skuteczności**

    - Problem: nie wiadomo, czy playbook ma sens, czy spowalnia pracę.
    - Jak uniknąć: zdefiniuj KPI dla playbooka (czas od intake do decyzji, liczba wyjątków, liczba incydentów związanych z AI).

Część B — Playbook „AI Tool Selection”

1. Purpose & Scope

Cel:

- Zapewnić **spójny, audytowalny i pragmatyczny sposób wyboru technologii AI** dla inicjatyw w organizacji.
- Zminimalizować ryzyka regulacyjne, reputacyjne i operacyjne związane z AI.
- Przyspieszyć drogę od pomysłu do produkcyjnego rozwiązania poprzez Fast Path/Safe Path.

Zakres:

- Każdy nowy lub znacząco zmieniany **use-case AI**, który:
    - wykorzystuje LLM/GenAI, RAG, agentów, klasyczny ML, CV/OCR, systemy rekomendacyjne lub optymalizację;
    - działa na danych klientów, pracowników, partnerów lub danych wrażliwych;
    - ma wpływ na decyzje biznesowe, finansowe, HR lub operacje.

Poza zakresem (opcjonalnie Fast Path Lite):

- Indywidualne eksperymenty w odizolowanych środowiskach sandbox z danymi syntetycznymi lub fikcyjnymi, **bez danych produkcyjnych**.

2. Definitions

- **GenAI (Generative AI)** – modele generujące tekst, obrazy, kod na podstawie promptu.

- **Prompt-only GenAI** – wykorzystanie GenAI bez dostępu do danych firmowych (poza danymi przekazywanymi w promptach) do generowania/edytowania treści.
- **RAG (Retrieval-Augmented Generation)** – architektura, w której GenAI korzysta z wektorowej bazy dokumentów/rekordów, aby odpowiadać na pytania i generować treści na bazie aktualnej wiedzy firmy.
- **Agent / agent z narzędziami** – system, w którym LLM nie tylko generuje tekst, lecz także planuje i wykonuje akcje w systemach (np. ERP/CRM) przy użyciu narzędzi/API.
- **Klasyczny ML** – modele regresji, klasyfikacji, rankingów, systemy rekomendacyjne, modele szeregów czasowych, anomaly detection, bez używania LLM jako głównego silnika.
- **CV/OCR (Computer Vision / Optical Character Recognition)** – modele rozpoznające obiekty/cechy na obrazach lub wyciągające tekst/struktury z dokumentów.
- **Optimization** – modele i algorytmy optymalizujące planowanie, przydział zasobów, routing (LP/MIP/heurystyki).
- **Rules/Automation** – deterministyczne reguły biznesowe, workflow engine, schedulery, bez komponentu uczącego się.
- **Fine-tuning** – dodatkowe trenowanie istniejącego LLM na danych domenowych.
- **HITL (Human-in-the-Loop)** – wymóg udziału człowieka w podejmowaniu decyzji, przeglądzie wyników lub zatwierdzeniu akcji.

3. When to use the Playbook (trigger conditions)

Playbook **musi być użyty**, gdy:

- Use-case obejmuje **LLM/GenAI** (w tym RAG, agenci) korzystające z **danych firmowych** lub działające na danych klientów/pracowników.
- Projekt zakłada **trenowanie modelu ML** na danych produkcyjnych (predykcja popytu, scoring, rekomendacje, anomalie).
- Rozwiązanie AI ma **wpływ na decyzje finansowe, HR, compliance, bezpieczeństwo** lub inne obszary high-risk opisane w EU AI Act.

Playbook **zalecany**, gdy:

- Zespół chce zbudować **powtarzalny wzorzec** (np. asystent RAG dla wewnętrznych zespołów).

Playbook **nie jest wymagany**, gdy:

- Eksperymenty w pełni offline / z danymi syntetycznymi, przeprowadzone przez pojedynczą osobę, bez integracji z systemami produkcyjnymi; wymaga jednak podlegania ogólnym politykom danych (brak uploadu danych poufnych do narzędzi osobistych).

4. Roles & Responsibilities (RACI)

Role:

- **Business Owner (BO)** – właściciel procesu i P&L.
- **Product Owner (PO)** – odpowiedzialny za wymagania i backlog rozwiązania AI.
- **Data/AI Lead (DAI)** – odpowiedzialny za wybór podejścia AI, architekturę modeli, eval.
- **IT/Architecture (ITA)** – odpowiedzialny za integracje, skalowalność, zgodność z architekturą.
- **Security (SEC)** – odpowiedzialny za bezpieczeństwo informacji, dostęp, logging.
- **Legal/Compliance (LEG)** – odpowiedzialny za zgodność z regulacjami (RODO, EU AI Act, sektorowe).
- **Operations/Run (OPS)** – odpowiedzialny za utrzymanie, monitoring, SLO.
- **QA/Test (QA)** – odpowiedzialny za testy funkcjonalne i niefunkcjonalne.
- **AI Governance Lead (AIG)** – właściciel playbooka, koordynuje governance gates.

RACI (wysoki poziom):

Etap	**BO**	**PO**	**DAI**	**ITA**	**SEC**	**LEG**	**OPS**	**QA**	**AIG**
Intake & triage	A	R	C	C	C	C	I	I	C
Wybór podejścia (Tool Selector)	A	R	R	C	C	C	I	I	C
MVI/MVP blueprint	A	R	R	R	C	C	C	C	C
Risk & controls	A	C	R	C	R	R	C	C	R
Implementacja	C	R	R	R	C	C	R	R	I

Etap	BO	PO	DAI	ITA	SEC	LEG	OPS	QA	AIG
Release + monitoring	A	R	C	C	C	C	R	C	C
Retrospektywa	A	R	C	C	C	C	C	C	R

5. Decision Workflow (end-to-end)

5.1 Intake (zgłoszenie potrzeby)

Wejście: AI Use Case Intake Form (szablon w Appendix).

Kroki:

1. BO/PO opisuje use-case (proces, job to be done, dane, oczekiwane KPI).
2. Formularz trafia do kolejki AI Governance (AIG) – np. jako ticket w Jira / ServiceNow.

Wyjście: zaakceptowany lub odrzucony intake (z komentarzami).

5.2 Triage (szybka kwalifikacja)

Cel: przypisać sprawę do **Fast Path** lub **Safe Path**.

Kryteria (przykładowe):

- **Typ danych:**
  - Niska wrażliwość – dane publiczne, syntetyczne.
  - Średnia – dane operacyjne bez danych osobowych.
  - Wysoka – dane osobowe, dane finansowe, dane zdrowotne, dane wrażliwe.
- **Wpływ na decyzje:**
  - Niski – generowanie wewnętrznych treści, asystent dla pracownika.
  - Średni – rekomendacje wspierające decyzje.
  - Wysoki – automatyczne decyzje finansowe/HR, wpływ na klientów.
- **Rodzaj komponentu AI:** LLM z RAG/agent, klasyczny ML, tylko reguły.

Reguła przykładowa:

- Jeśli (wysoka wrażliwość danych **lub** wysoki wpływ na decyzje **lub** użycie agenta z write-access) → **Safe Path**.

- W przeciwnym razie → **Fast Path**.

5.3 Wybór podejścia (na bazie AI Tool Selector)

W tym etapie zespół (BO, PO, DAI, ITA) przechodzi wspólnie przez **AI Tool Selector**.

Streszczenie logiki AI Tool Selector:

1. Jakie są główne dane wejściowe? (tekst, dane strukturalne, obrazy/skany).
2. Jaki jest główny typ zadania? (generowanie, streszczanie, ekstrakcja/klasyfikacja, Q&A, rekomendacje, predykcja, anomalie, optymalizacja, prosta automatyzacja, multi-step workflow).
3. Jak wysoka jest stawka błędu i czy potrzebna jest ścisła zgodność z dokumentami / cytaty?
4. Czy AI ma wykonywać akcje w systemach (agent), czy tylko sugerować?
5. Na tej podstawie wybierany jest typ rozwiązania: Prompt-only GenAI, RAG, Agent, klasyczny ML, OCR/CV, Optimization, Rules/Automation.

Instrukcja użycia na warsztacie:

- Użyj **AI Tool Selection Worksheet** (Appendix) jako „ścieżki po diagramie”.
- PO czyta kolejne pytania, DAI/ITA pomagają zaklasyfikować.
- Każda odpowiedź zawęża wybór podejścia.
- Na końcu zespół zapisuje wybór w **Decision Record**.

5.4 MVI/MVP blueprint

Dla wybranego podejścia AI zespół definiuje **Minimal Viable Implementation (MVI)**:

- Zakres funkcjonalny (co robi MVI, czego *nie* robi).
- Dane wejściowe i wyjściowe.
- Minimalna architektura (np. RAG: wektorowa baza, retriever, LLM; ML: feature store, model, batch scoring).
- Wymagany HITL dla decyzji high-risk.
- Plan testów (offline eval + smoke tests).

5.5 Risk & Controls

Na tym etapie wykonywana jest **Risk Assessment Checklist (LLM/RAG/Agent/ML)**.

Kontrole typowe:

- Klasyfikacja ryzyka według EU AI Act (unacceptable / high / limited / minimal).
- Mapowanie na NIST AI RMF (Govern/Map/Measure/Manage) – czy mamy polityki, kontekst, metryki, plany zarządzania ryzykiem.

- Kontrole ISO/IEC 42001 – inventory systemów AI, role, procedury, PDCA.
- Kontrole specyficzne dla LLM/RAG/agentów (halucynacje, grounding, data leakage, misuse tools).

Wynikiem jest **Risk & Controls Summary** (część Decision Record) oraz przypisanie use-case'u do Fast/Safe Path (lub eskalacja).

5.6 Implementacja i testy

- ITA/DAI realizują implementację zgodnie z MVI.
- QA przygotowuje **Evaluation Plan** (offline + online).
- Wykonywane są:
    - Testy funkcjonalne i niefunkcjonalne.
    - Testy LLM/RAG (patrz sekcja Quality & Evaluation).
    - Testy bezpieczeństwa (SEC) i privacy (LEG/SEC).

5.7 Release + Monitoring

- Przejście przez **Go/No-Go Checklist**.
- Dla komponentów high-risk – decyzja release'u na forum komitetu AI Governance.
- Po wdrożeniu – włączenie monitoringu:
    - metryki techniczne (quality, latency, cost);
    - metryki biznesowe (KPI procesu);
    - alerty bezpieczeństwa i anomalie.

5.8 Retrospektywa i iteracja

Po 4–12 tygodniach od wdrożenia:

- BO/PO, DAI, OPS, AIG organizują **Post-Implementation Review**.
- Ocena ROI, adoption, incydentów, jakości modeli.
- Wnioski wpływają na:
    - aktualizację modeli/promptów;
    - aktualizację playbooka lub Tool Selectora (przez Playbook Update Request).

6. Governance Gates

W zależności od klasy ryzyka ustala się następujące „bramki" (gates):

6.1 Fast Path (niska/średnia stawka błędu)

Przykłady: wewnętrzne generowanie treści, asystent dla pracownika, rekomendacje niskiego ryzyka.

Wymagane gates:

1. **Gate F1 – Intake & triage**

    - Wypełniony Intake Form.
    - Klasyfikacja jako Fast Path (potwierdzona przez AIG/SEC).

2. **Gate F2 – Tool Selection & MVI**

    - Uzupełniony Tool Selection Worksheet.
    - Decision Record zaakceptowany przez BO i DAI.

3. **Gate F3 – Minimalne testy**

    - Próbka testów offline (np. 20–50 przykładów); akceptowalny poziom jakości.
    - Sprawdzenie, że system nie loguje danych wrażliwych poza kontrolą.

4. **Gate F4 – Release**

    - Podpis BO + PO + DAI.
    - Rejestracja systemu AI w inventory.

6.2 Safe Path (wysoka stawka błędu)

Przykłady: scoring kredytowy, HR screening, automatyczne decyzje finansowe, działania agentów z write-access do systemów core.

Wymagane gates:

1. **Gate S1 – Intake & triage (rozszerzony)**

    - Klasyfikacja jako high-risk AI.
    - Wstępna analiza wpływu (impact assessment).

2. **Gate S2 – Tool Selection & MVI**

    - Warsztat z pełnym składem RACI.
    - Decision Record zatwierdzony przez BO, DAI, ITA, SEC, LEG.

3. **Gate S3 – Risk & Controls Approval**

    - Risk Assessment Checklist (LLM/RAG/Agent/ML) zakończona.
    - Mapa ryzyk i kontrole zgodnie z NIST AI RMF i ISO 42001.
    - Decyzja komitetu AI Governance (AIG + przedstawiciele funkcji).

4. **Gate S4 – Evaluation & Security Review**

    - Pełen Evaluation Plan z wynikami offline.
    - Testy bezpieczeństwa (penetracyjne, data leakage, abuse).
    - Legal review (zgodność z regulacjami sektorowymi i EU AI Act).

5. **Gate S5 – Go/No-Go & Controlled Release**

    - Go/No-Go Checklist spełniona.
    - Pilotaż w ograniczonym zakresie (np. 5–10% ruchu, shadow mode dla agentów).
    - Formalny podpis: BO, CIO, CISO, CCO/GC.

6. **Gate S6 – Post-Market Monitoring**

    - Plan monitoringu i post-market surveillance zgodny z ISO/IEC 42001 i EU AI Act.

7. Quality & Evaluation

7.1 Metryki techniczne

- **LLM / GenAI / RAG**

    - Accuracy / Precision / Recall na zbiorze eval.
    - Hallucination rate (procent odpowiedzi niezgodnych z ground-truth lub poza zakresem).
    - Groundedness score (odsetek odpowiedzi powołujących się na zindeksowane źródła dla RAG).
    - Latency (P50, P95, P99) i cost per call.

- **Klasyczny ML / forecasting / anomaly detection**

    - Accuracy, AUC, precision/recall, ROC curves.
    - MAE/RMSE/MAPE dla forecasting.
    - TPR/FPR dla detekcji anomalii.

7.2 Metryki biznesowe

- Czas obsługi procesu (cycle time).
- Koszt jednostkowy procesu.
- FCR / NPS / CSAT (w obsłudze klienta).
- Liczba incydentów związanych z błędami AI.
- ROI (oszczędności vs koszty wdrożenia).

7.3 Eval dla LLM/RAG/agentów

- **Offline eval:**
    - Zbiór referencyjny (prompt → oczekiwana odpowiedź / etykieta).
    - Testy halucynacji (pytania spoza wiedzy lub z konfliktami).
    - Testy bezpieczeństwa (prompt injection, jailbreak, wycieki danych).
- **Online monitoring:**
    - Sampling odpowiedzi i ręczny review (HITL sampling).
    - Feedback użytkowników (thumbs up/down, komentarze).
    - Alerty na skoki w metrykach (spadek jakości, wzrost latency/kosztu).
- **Testy regresji promptów:**
    - Utrzymanie zestawu testów regression suite; każda zmiana promptu/policy jest testowana offline.
- **Testy retrieval (RAG):**
    - Coverage (czy odpowiednie dokumenty są w top-K wyników).
    - Precision@K dla dokumentów.
- **Testy narzędzi (tools/agents):**
    - Symulacje akcji agenta w sandboxie.
    - Warunki bezpieczeństwa (max liczba akcji, limity operacji, listy dozwolonych/zakazanych tooli).

8. Security / Privacy / IP – minimalny zestaw zasad

- **Zakaz shadow AI** – zabronione jest używanie prywatnych kont narzędzi AI do danych firmowych; wszystkie rozwiązania muszą korzystać z zatwierdzonych platform.
- **Data minimization** – przekazujemy do modeli tylko dane niezbędne; domyślnie pseudonimizacja/anonymizacja.
- **Access control & logging** – role-based access, audyt logów promptów, odpowiedzi, użytych narzędzi.
- **Data residency & retention** – wymagania dot. lokalizacji danych i czasu przechowywania zgodne z regulacjami i politykami.
- **IP & licensing** – zasady używania generowanych treści (np. ograniczenia co do włączania treści do produktów), weryfikacja licencji modeli.

9. Change Management & Versioning

- Playbook ma wersjonowanie semantyczne (np. v1.2.0).
- Zmiany inicjowane poprzez **Playbook Update Request**.
- Istotne zmiany wymagają przeglądu przez AI Governance Committee.
- Każda zmiana jest komunikowana (newsletter, changelog w wiki, krótkie szkolenie).

10. Appendix – instrukcja użycia AI Tool Selector

- **Krok 1:** Zespół zbiera się na 60–90 min z wypełnionym Intake Form.
- **Krok 2:** Wspólnie przechodzicie przez AI Tool Selection Worksheet (od pytań o dane, przez typ zadania, po stawkę błędu i autonomię).
- **Krok 3:** DAI/ITA upewniają się, że odpowiedzi są spójne z architekturą danych.
- **Krok 4:** Wybierane jest podejście AI, co jest dokumentowane w Decision Record.
- **Krok 5:** Na tej podstawie powstaje MVI blueprint.

Templates (Markdown)

Poniższe szablony są gotowe do skopiowania do wiki / narzędzi.

1. AI Use Case Intake Form

```
# AI Use Case Intake Form

## 1. Podstawowe informacje

- Nazwa use-case'u:

- Data zgłoszenia:

- Zgłaszający (imię, rola, dział):

- Business Owner:

- Product Owner:

## 2. Opis procesu i celu

- Opis procesu (1–3 akapity):

- „Job to be done” (co ma być lepiej / taniej / szybciej?):
```

- Użytkownicy docelowi (rola, liczba):

3. Dane

- Jakie dane będą używane? (typy, moduły/systemy źródłowe):

- Czy dane zawierają dane osobowe / wrażliwe? (tak/nie + opis):

- Czy dane pochodzą z systemów regulowanych (finanse, zdrowie, krytyczna infrastruktura)?

4. Wpływ i ryzyko

- Jaki jest wpływ rozwiązania na klientów/pracowników? (niski/średni/wysoki + opis):

- Czy AI będzie **podejmować decyzje**, czy tylko **sugerować**?:

- Przykłady potencjalnych błędów i ich konsekwencje:

5. KPI i sukces

- Docelowe KPI (biznesowe):

- Docelowe KPI (techniczne):

- Horyzont czasowy (MVP, pełne wdrożenie):

6. Inne

- Powiązane inicjatywy/projekty:

- Dodatkowe komentarze:

2. AI Tool Selection Worksheet

AI Tool Selection Worksheet

1. Dane wejściowe

1. Główny typ danych:

   - [ ] Tekst / dokumenty

- [ ] Dane strukturalne (tabele, logi, eventy)

- [ ] Obrazy / skany / wideo

2. Główne systemy źródłowe (ERP/CRM/HR/…):

2. Typ zadania

3. Główny typ problemu:

- [ ] Generowanie treści

- [ ] Streszczanie

- [ ] Ekstrakcja / klasyfikacja

- [ ] Wyszukiwanie wiedzy / Q&A

- [ ] Rekomendacje

- [ ] Predykcja / forecasting

- [ ] Wykrywanie anomalii

- [ ] Planowanie / optymalizacja

- [ ] Prosta automatyzacja

- [ ] Multi-step workflow / agent

3. Ryzyko i autonomia

4. Stawka błędu:

- [ ] Niska

- [ ] Średnia

- [ ] Wysoka

5. Czy wymagane są cytaty / źródła (grounding)? (tak/nie)

6. Czy AI ma wykonywać akcje w systemach (np. tworzyć/edytować rekordy)?

- [ ] Nie, tylko sugestie

- [ ] Tak, z zatwierdzeniem człowieka

- [ ] Tak, automatycznie (opis warunków)

4. Wybór podejścia (na bazie diagramu)

Na podstawie odpowiedzi i diagramu AI Tool Selector wybierz podejście (zaznacz jedno główne + ewentualnie pomocnicze):

- [ ] Prompt-only GenAI

- [ ] RAG (LLM + wektorowa baza)

- [ ] Agent z narzędziami

- [ ] Klasyczny ML (klasyfikacja/regresja)

- [ ] Time-series forecasting

- [ ] Anomaly detection

- [ ] Optimization

- [ ] CV/OCR / Document AI

- [ ] Rules / Automation

Uzasadnienie wyboru (1–3 akapity):

Alternatywy rozważane i odrzucone (z krótkim uzasadnieniem):

5. MVI (Minimal Viable Implementation)

- Zakres funkcji MVI:

- Dane wejściowe/wyjściowe:

- Wstępny plan testów:

3. Risk Assessment Checklist (LLM/RAG/Agent/ML)

Risk Assessment Checklist (LLM / RAG / Agent / ML)

1. Klasyfikacja ryzyka

- Kategoria wg EU AI Act:
  - [ ] Minimal risk
  - [ ] Limited risk
  - [ ] High risk
  - [ ] Unacceptable (projekt wstrzymany)
- Typ danych:
  - [ ] Publiczne
  - [ ] Wewnętrzne
  - [ ] Dane osobowe
  - [ ] Dane wrażliwe / szczególne kategorie

2. LLM / RAG / Agent – specyficzne ryzyka

- Czy model może halucynować treści o wysokiej stawce błędu? (tak/nie + opis)
- Czy odpowiedzi muszą być oparte na konkretnych dokumentach? (tak/nie)
- Czy zastosowano RAG z cytatami dla krytycznych odpowiedzi? (tak/nie)
- Czy istnieje ryzyko wycieku danych do zewnętrznych dostawców modeli? (opis)
- Czy agent ma write-access do systemów core? (tak/nie + zakres)
- Czy wdrożono HITL dla high-risk kroków? (tak/nie)

3. Zarządzanie ryzykiem (NIST AI RMF / ISO 42001)

- Czy use-case jest zarejestrowany w inventory AI systems? (tak/nie)

- Czy zidentyfikowano interesariuszy i kontekst (Map)? (tak/nie)

- Czy zdefiniowano metryki jakości i ryzyka (Measure)? (tak/nie)

- Czy zaplanowano monitoring i działania korygujące (Manage)? (tak/nie)

4. Decyzja

- Podsumowanie głównych ryzyk:

- Zastosowane kontrole:

- Rekomendacja:

  - [ ] Kontynuować (Fast Path)

  - [ ] Kontynuować (Safe Path)

  - [ ] Wstrzymać / eskalować

Podpisy: BO / DAI / SEC / LEG / AIG

4. Evaluation Plan Template (offline + online)

AI Evaluation Plan

1. Zakres

- Use-case:

- Komponenty do ewaluacji (LLM/RAG/ML/Agent/…):

2. Offline eval

- Zbiory testowe:

  - Opis datasetów (wielkość, pochodzenie, metody anonimizacji):

- Metryki:

  - Techniczne: (np. accuracy, F1, MAE, hallucination rate, groundedness score):

  - Progi akceptacji:

- Procedura testowania:

  - Jak często:

  - Kto odpowiada:

3. Online monitoring

- Metryki monitorowane w produkcji:

- Alert thresholds (np. spadek jakości > X%, wzrost latency > Y%):

- Sampling do HITL review (np. 5% odpowiedzi tygodniowo):

- Kanały zgłaszania issue (np. przycisk feedback w UI):

4. Raportowanie

- Częstotliwość raportów (np. miesięcznie):

- Odbiorcy (BO, PO, DAI, OPS, AIG):

5. Go/No-Go Checklist

Go/No-Go Checklist (AI Release)

1. Funkcjonalność i jakość

- [ ] Zrealizowano wymagania MVI/MVP.

- [ ] Offline eval spełnia ustalone progi.

- [ ] Przeprowadzono testy regresji promptów / modeli.

2. Bezpieczeństwo i prywatność

- [ ] Przeprowadzono przegląd bezpieczeństwa (SEC).

- [ ] Dane są klasyfikowane i zabezpieczone zgodnie z politykami.

- [ ] Brak wycieków danych do niezatwierdzonych dostawców.

3. Governance i ryzyko

- [ ] Risk Assessment Checklist ukończony.

- [ ] System wpisany do inventory AI.

- [ ] Dla high-risk: zatwierdzenie komitetu AI Governance.

4. Operacje

- [ ] Monitoring i alerting skonfigurowane.

- [ ] Runbooki operacyjne przygotowane (incident response, rollback).

- [ ] Wdrożony mechanizm wyłączenia / rollbacku.

Decyzja:

- [ ] Go

- [ ] No-Go (z komentarzem):

Podpisy: BO / CIO / CISO / AIG / inni:

6. Post-Implementation Review

Post-Implementation Review (AI Solution)

1. Podsumowanie

- Nazwa rozwiązania:

- Data wdrożenia:

- Zakres MVI / aktualny zakres:

2. Wyniki biznesowe

- KPI bazowe vs aktualne (tabela):

- ROI (oszacowanie):

- Opinie użytkowników (NPS / jakościowe):

3. Wyniki techniczne

- Zmiany metryk jakości modeli:

- Incydenty / błędy / near-misses:

4. Wnioski

- Co działa dobrze:

- Co nie działa / co zaskoczyło:

- Rekomendacje na kolejną iterację:

5. Wpływ na playbook

- Czy playbook wymaga aktualizacji? (tak/nie)

- Opis proponowanych zmian:

7. Playbook Update Request

Playbook Update Request – AI Tool Selection

- Zgłaszający (imię, rola, dział):

- Data:

1. Obszar playbooka

- [ ] Purpose & Scope

- [ ] Definitions

- [ ] Triggers

- [ ] RACI

- [ ] Decision Workflow

- [ ] Governance Gates

- [ ] Quality & Evaluation

- [ ] Security / Privacy / IP

- [ ] Templates

- [ ] Inne:

2. Opis zmiany

- Aktualny stan (cytat / skrót):

- Proponowana zmiana:

- Uzasadnienie (doświadczenia, incydenty, nowe regulacje):

3. Wpływ

- Wpływ na zespoły/procesy:

- Pilność (niska/średnia/wysoka):

4. Załączniki

- Linki do decyzji, incydentów, raportów:

Sekcja dla AI Governance Lead:

- Decyzja:
  - [ ] Zaakceptowano
  - [ ] Odrzucono
  - [ ] Wymaga doprecyzowania
- Plan wdrożenia zmiany (wersja, data, komunikacja):