AI W USŁUGACH PŁATNICZYCH. AUTOMATYZACJA, PERSONALIZACJA I WYZWANIA REGULACYJNE W ŚWIETLE AI ACT.

Sygn. PAB/WIB 17/2025

ISBN: 978-83-976731-2-0

Raport opracowany na zlecenie Programu Analityczno-Badawczego Fundacji Warszawski Instytut Bankowości.

Warszawa, sierpień 2025

PROGRAM ANALITYCZNO BADAWCZY

O raporcie

Raport „AI w usługach płatniczych. Automatyzacja, personalizacja i wyzwania regulacyjne w świetle AI Act." został opracowany na zlecenie Programu Analityczno-Badawczego Fundacji Warszawski Instytut Bankowości.

Autorzy

Raport przygotowany przez zespół w składzie:

Dariusz Szostek

CEO Szostek_Digital i Wspólnicy. Profesor Wydziału Prawa i Administracji Uniwersytetu Śląskiego, radca prawny Specjalista z zakresu paperless, kwalifikowanych usług zaufanych, tożsamości cyfrowej, cyfrowego obiegu dokumentów, algorytmizacji procesów, implementacji w organizacji algorytmów w tym AI w zgodności z wymogami cyberbezpieczeństwa, tokenizacji procesów w tym wykorzystanie blockchain oraz smart contract, implementacji prawa w algorytmy (inżynieria prawa), cywilista. Realizator projektów Horizon 2020 związanych z innowacjami w przemyśle – SHOP4CF i MAS4AI. Współtwórca m.in.: koncepcji i legislacji eSądu w Lublinie – elektronicznego postępowania upominawczego; koncepcji i legislacji elektronicznego potwierdzenia odbioru, powszechnie stosowanego na Poczcie Polskiej (tablety na których potwierdza się odbiór pism); koncepcji i wdrożenia podpisu biometrycznego na tabletach min. biocertiX (wspólnego projektu Samsunga, Asseco i Xtension), koncepcji i wdrożenia oprogramowania Ius Case w wydawnictwie C.H. Beck. Założyciel i dyrektor Centrum naukowego – CYBER SCIENCE – Śląskiego Centrum Inżynierii Prawa, Technologii i Kompetencji Cyfrowych, zrzeszającego Uniwersytet Śląski, Politechnikę Śląską, Uniwersytet Ekonomiczny w Katowicach, Instytut EMAG sieci Łukasiewicza i NASK. Działacz w zakresie zarządzania Internetem, członek rady programowej IGF Poland Ekspert kadencji 2020–2024 Obserwatorium Parlamentu Europejskiego ds. Sztucznej inteligencji. Przewodniczący Zespołu ds. Cyberbezpieczeństwa Konferencji Rektorów Akademickich Szkół Polskich. W 2025 r. powołany przez wicepremiera K. Gawkowskiego na członka Rady Naukowej Instytutu Łączności Autor licznych książek i opracowań w zakresie prawa nowych technologii – „Legal tech. Czyli jak bezpiecznie korzystać z narzędzi informatycznych w organizacji, w tym w kancelarii oraz dziale prawnym". Inicjator serii metodyk związanych z technologią i prawem w wydawnictwie C.H. Beck. Członek licznych projektów w zakresie cyfryzacji. Popularyzator wiedzy w zakresie Blockchain – członek Blockhaton EUiPO oraz autor między innymi Blockchain i Prawo (w wersji angielskiej Blockchain and law wyd. Nomos). Współtwórca i partner w latach 2014–2024 kancelarii Szostek_Bar i Partnerzy, w latach 2017–2019 współpracownik PwC Legal, w latach 2019–2021 współpracownik Kancelarii Maruta-Wachta. Odpowiedzialny za szereg wdrożeń paperless w organizacjach (m.in. w bankach). Wykładowca akademicki – profesor na WPIA UŚ, ale także wykładowca studiów podyplomowych z zakresu nowych technologii, AI czy też cyberbezpieczeństwa na: Akademii Koźmińskiego, Polskiej Akademii Nauk, Politechnice Śląskiej (Zarządzanie cyberbezpieczeństwem).

Rafał Prabucki

COO Szostek _Digital i Wspólnicy. Doktor nauk prawnych i inżynier. Auditor wiodący ISO/IEC 27001, BCMS ISO 22301 oraz ISO/IEC 42001/2023. Posiadacz tytułu Certified in Cybersecurity (CC) wydanym przez (ISC)². Adiunkt na Uniwersytecie Śląskim. Członek CYBER SCIENCE i SABI. Założyciel Legal-Hackers Katowice. Asystent w zakończonych projektach dotyczących wykorzystania nowych technologii w przemyśle: MAS4AI na Uniwersytecie Śląskim i SHOP4CF na Uniwersytecie Opolskim. Alumn stypendiów w Polsce, Hiszpani i Niemczech. Prowadzący wykłady na Uczelniach w Polsce, Litwie i we Francji. Członek Społecznego Zespołu Ekspertów przy Prezesie Urzędu ochrony danych Osobowych. Autor książek i komentarzy, uczestnik Hackathonów i konkursów na innowacje – finalista w Młodzi i Innowacyjni dla PGNiG w 2017 roku oraz drugie miejsce w #hack4law w edycji z 2024 roku.

O programie

Program Analityczno-Badawczy przy Fundacji WIB powstał w 2019 r. jako odpowiedź na potrzeby sektora bankowego w zakresie wysokiej jakości analiz i badań z obszaru cyberbezpieczeństwa i nowych technologii, a także szeroko rozumianego otoczenia sektora bankowego, kształtującego warunki działania banków w Polsce. Prace analityczno-badawcze w ramach programu realizowane są pod kątem możliwości praktycznego wy korzystania wyników w celu rozwoju sektora bankowego, podnoszenia poziomu bezpieczeństwa usług oraz – w ostateczności – kreowania wartości dla klientów – końcowych użytkowników bankowości.

W ramach programu realizowane są analizy i badania w następujących obszarach:

- Nowe technologie w sektorze bankowym, cyberbezpieczeństwo banków i systemy płatnicze
- Ekonomiczna analiza rozwoju sektora bankowego, stabilności banków, otoczenia makroekonomicznego
- Prawno-regulacyjne i legislacyjne uwarunkowania działalności sektora bankowego oraz wyzwania wynikające z orzecznictwa sądów krajowych i europejskich
- ESG i sustainable finance
- Prace analityczno-badawcze w zakresie poszczególnych obszarów funkcjonowania banków oraz w zakresie funkcjonowania wskaźników referencyjnych

Więcej na temat działalności programu na stronie www.pabwib.pl

Kontakt

dr Tomasz Pawlonka
Dyrektor Programu
m: 505 917 778
e: tpawlonka@wib.org.pl

Warszawski Instytut Bankowości
00-380 Warszawa, ul. Kruczkowskiego 8
t: (22) 182 31 70
e: pab@wib.org.pl

Agnieszka Nierodka
m: 607 484 391
e: anierodka@wib.org.pl

dr Agnieszka Wicha
m: 661 646 474
e: agnieszka.wicha@zbp.pl

Bartosz Przyborowski
m: 795 933 104
e: bartosz.przyborowski@zbp.pl

Spis treści

1. Wprowadzenie

Współczesna gospodarka cyfrowa nieustannie przekształca się pod wpływem rozwoju zaawansowanych technologii, spośród których szczególne miejsce zajmuje sztuczna inteligencja (AI – Artificial Intelligence). Zastosowanie AI znajduje dziś coraz szersze zastosowanie w sektorach o kluczowym znaczeniu społecznym i gospodarczym, takich jak opieka zdrowotna, administracja publiczna, transport, edukacja czy – co szczególnie istotne w kontekście niniejszego opracowania – sektor finansowy. Rosnące zainteresowanie tymi technologiami wiąże się z obietnicą poprawy efektywności operacyjnej, lepszego zarządzania ryzykiem, automatyzacji procesów i podniesienia jakości obsługi klienta. Jednak równocześnie rośnie świadomość wyzwań, jakie niesie ze sobą ich wdrożenie – zarówno na płaszczyźnie technicznej, jak i prawno-etycznej.

Systemy AI mają coraz większy wpływ na podejmowanie decyzji – nie tylko wspomagają, ale często też zastępują człowieka w procesach analitycznych, prognostycznych i decyzyjnych. Oznacza to, że ich działanie musi być przedmiotem szczególnej kontroli i nadzoru. Decyzje podejmowane przez algorytmy mogą bowiem prowadzić do konsekwencji o dużej wadze – np. odmowy kredytu, wadliwej ocenie ryzyka klienta, zablokowania transakcji czy wykrycia rzekomego oszustwa. Każda z tych decyzji musi być nie tylko technicznie poprawna, ale także przejrzysta, sprawiedliwa i zgodna z obowiązującym prawem. To właśnie w tym kontekście pojawia się potrzeba kompleksowego systemu **monitorowania, oceny i raportowania działania systemów sztucznej inteligencji**.

Rozporządzenie AI Act, przyjęte przez Unię Europejską, wyznacza nowe ramy regulacyjne dla rozwoju i stosowania AI, ze szczególnym naciskiem na tzw. systemy wysokiego ryzyka. Wśród nich centralne miejsce zajmują rozwiązania stosowane w sektorze finansowym. AI Act nie jest jednakże jedyną regulacją nakładającą na instytucje finansowe nowe obowiązki. Uzupełnia je szereg regulacji z zakresu danych, w tym danych osobowych oraz cyberbezpieczeństwa na czele z DORA. Zgodnie z AI Act, dostawcy i użytkownicy AI zobowiązani są do spełnienia szeregu wymogów – w tym prowadzenia systematycznego nadzoru nad funkcjonowaniem systemu, zapewnienia możliwości interwencji człowieka, przeprowadzania audytów, testów odporności oraz raportowania incydentów. Przepisy te stanowią odpowiedź na liczne zagrożenia związane z AI, w tym niezamierzoną dyskryminację, brak wyjaśnialności decyzji, podatność na manipulacje oraz ryzyko destabilizacji procesów finansowych.

W odpowiedzi na te regulacje oraz realne wyzwania praktyczne, organizacje wdrażające AI – a w szczególności instytucje finansowe – muszą przyjąć podejście systemowe i wielowymiarowe. Obejmuje ono zarówno narzędzia techniczne jak i procedury organizacyjne (audyt wewnętrzny, struktura QMS, lista kontrolna zgodności). Kluczową rolę odgrywa również **monitorowanie post-market**, czyli kontrola działania systemu po jego wdrożeniu – w realnych warunkach funkcjonowania, na żywych danych, z uwzględnieniem zmiennych scenariuszy i zachowań użytkowników.

Jednym z najważniejszych aspektów nadzoru nad AI jest zapewnienie **transparentności** – zarówno wobec użytkownika końcowego, jak i wobec organów nadzoru. Transparentność oznacza nie tylko możliwość poznania podstawy danej decyzji, ale także zapewnienie rozliczalności, identyfikowalności i komunikowalności systemu. W kontekście sektora finansowego, transparentność zyskuje dodatkowy wymiar – wpływa bowiem bezpośrednio na zaufanie klientów do instytucji, które podejmują decyzje w sposób częściowo zautomatyzowany.

Istotnym zagadnieniem w kontekście raportowania AI jest również **wyjaśnialność (explainability)** – zdolność systemu do uzasadnienia swoich decyzji w sposób zrozumiały dla człowieka. Mechanizmy explainable AI pozwalają nie tylko zrozumieć, ale także weryfikować trafność i sprawiedliwość decyzji. Ich implementacja jest szczególnie istotna w środowiskach wymagających wysokiego poziomu odpowiedzialności regulacyjnej, takich jak bankowość i ubezpieczenia.

Nie sposób pominąć również kwestii **uprzedzeń algorytmicznych (bias)**. Systemy AI – jeśli nie są odpowiednio testowane i monitorowane – mogą powielać, a nawet pogłębiać istniejące nierówności społeczne. Problem ten dotyczy nie tylko danych treningowych, ale również konstrukcji modelu oraz polityki decyzyjnej wdrażanej przez organizację. Z tego względu dokumentowanie i testowanie na obecność biasu – za-

równo na poziomie danych, jak i wyników – staje się integralnym elementem zgodnego z prawem i etyką wdrażania AI.

Na koniec warto podkreślić, że skuteczne raportowanie nie kończy się na analizie technicznej. Jego istotą jest także tworzenie dokumentacji, która pozwala na **niezależną ocenę systemu** – zarówno przez organy nadzoru (np. KNF, EBA), jak i przez wewnętrzne zespoły audytu. Dokumentacja ta powinna obejmować wyniki testów funkcjonalnych, analizę ryzyka, oceny zgodności z AI Act oraz rekomendacje dotyczące dalszego doskonalenia modelu. Praktyka ta wzmacnia rozliczalność organizacji i stanowi podstawę zaufania społecznego do technologii.

Podsumowując, wprowadzenie systemów sztucznej inteligencji do sektora finansowego wymaga nie tylko kompetencji technicznych, ale przede wszystkim **systemowego podejścia do oceny, nadzoru i raportowania ich działania**. Dokument niniejszy stanowi próbę zarysowania najlepszych praktyk, narzędzi oraz ram regulacyjnych w tym zakresie, z uwzględnieniem wyzwań i zaleceń wynikających z AI Act oraz praktyk krajowych i międzynarodowych.

dr hab. prof. UŚ Dariusz Szostek

dr inż. Rafał Prabucki

1.1. Cel i zakres raportu

Celem niniejszego raportu jest kompleksowa analiza zastosowania systemów sztucznej inteligencji (AI) w sektorze usług płatniczych, ze szczególnym uwzględnieniem wymogów prawnych wynikających z rozporządzenia AI Act oraz powiązanych regulacji. Raport ma na celu:

- zidentyfikowanie głównych obszarów zastosowania AI w usługach płatniczych: automatyzacji, personalizacji, zarządzania ryzykiem i obsługi klienta;
- przedstawienie zagrożeń wynikających z implementacji AI, w tym ryzyk systemowych, algorytmicznych i cybernetycznych;
- ocenę wpływu AI Act na projektowanie, wdrażanie, monitoring i audyt systemów AI w instytucjach finansowych;
- sformułowanie praktycznych rekomendacji dotyczących zgodności regulacyjnej, raportowania oraz nadzoru nad systemami AI klasyfikowanymi jako wysokiego ryzyka.

Raport skierowany jest zarówno do decydentów i menedżerów w sektorze finansowym, jak i do organów regulacyjnych, prawników technologicznych, specjalistów ds. compliance i inżynierów odpowiedzialnych za wdrażanie rozwiązań AI.

Zakres raportu obejmuje:

- analizę rzeczywistych wdrożeń AI w bankowości i fintechach,
- studia przypadków (case studies) wybranych incydentów i systemów AI (np. voice cloning, deepfake),
- ocenę zgodności modeli AI z wymaganiami prawnymi, technicznymi oraz etycznymi,
- przegląd metod nadzoru post-market oraz standardów raportowania zgodnego z AI Act,
- identyfikację dobrych praktyk zgodnych z normami ISO/IEC 42001, ISO/IEC 27001 i ISO/IEC 23894.

Raport został przygotowany w oparciu o podejście jakościowe (qualitative), z elementami analizy funkcjonalno-prawnej oraz studiów przypadku. Metodologia obejmowała:

- analizę prawną: w tym analizę tekstu AI Act oraz dokumentów powiązanych;
- analizę funkcjonalną systemów AI: w szczególności w kontekście ich wpływu na podejmowanie decyzji, zgodność z zasadami przejrzystości i wyjaśnialności, zarządzanie ryzykiem;
- desk research obejmujący literaturę międzynarodową, raporty think-tanków (Alan Turing Institute, CEPS), publikacje regulatorów oraz dokumentację techniczną standardów ISO;
- analizę przypadków użycia (use cases) oraz incydentów związanych z AI w sektorze finansowym.

Do opracowania raportu wykorzystano autorską metodykę oceny dojrzałości systemów AI w instytucjach finansowych, inspirowaną podejściem risk-based oraz zintegrowaną z cyklem życia systemu AI. Uwzględniono:

- cykl życia AI wg modelu D. Lesliego (The Alan Turing Institute),
- katalogi kontrolne AICRIV i ATRS,
- system QMS zgodny z art. 17 AI Act,
- podejście ENISA do wielowarstwowego bezpieczeństwa systemów AI,
- zastosowanie metryk wyjaśnialności (LIME, SHAP), odporności (F1-score, AUC), etyczności (bias scores) i rozliczalności (auditability indicators).

Oceniano m.in.:

- sposób zarządzania danymi treningowymi i operacyjnymi,
- wdrożenie nadzoru ludzkiego (human oversight),
- zgodność z zasadami FAIR i Explainable AI (XAI),
- mechanizmy reagowania na incydenty i prowadzenia post-market monitoring.

Raport ma charakter interdyscyplinarny – łączy perspektywę prawną, techniczną oraz etyczną w analizie systemów AI wykorzystywanych w usługach płatniczych.

1.2. Znaczenie sztucznej inteligencji w sektorze płatniczym

Sektor płatniczy, będący filarem współczesnych systemów finansowych, dynamicznie ewoluuje pod wpływem technologii sztucznej inteligencji (AI). Rozwiązania oparte na AI umożliwiają automatyzację i optymalizację procesów płatniczych, zwiększając ich bezpieczeństwo, efektywność oraz dostępność. Kluczowe zastosowania AI w tym obszarze obejmują wykrywanie oszustw, ocenę ryzyka transakcji, personalizację usług oraz poprawę doświadczeń klienta[1]. Na przykład AI umożliwia automatyzację zadań takich jak[2]:

- weryfikacja tożsamości użytkownika,
- wykrywanie podejrzanych transakcji,
- dynamiczne ustalanie limitów płatności,
- personalizacja oferty na podstawie historii transakcji.

AI wspiera też procesy innowacyjności. AI to rozwój[3]:

- płatności bezdotykowych i checkout-free (np. Amazon Go),
- rozpoznawania twarzy przy kasach,
- adaptacyjnych systemów płatniczych opartych na analizie kontekstu użytkownika.

Jednym z najbardziej przełomowych obszarów wykorzystania AI w płatnościach jest detekcja nadużyć finansowych. Systemy oparte na generatywnej AI, jak np. platforma Mastercard Decision Intelligence, analizują w czasie rzeczywistym ogromne wolumeny danych transakcyjnych i behawioralnych, identyfikując anomalie oraz wzorce potencjalnego oszustwa. Dzięki temu możliwe jest przewidywanie i zapobieganie oszustwom zanim zostaną one zrealizowane, co przekłada się na znaczący wzrost poziomu bezpieczeństwa użytkowników usług płatniczych[4]. W kontekście bezpieczeństwa możemy wyszczególnić kilka kluczowych obszarów[5]:

1. monitorowanie transakcji finansowych w czasie rzeczywistym – AI wspiera identyfikację podejrzanych transakcji finansowych, np. w ramach przeciwdziałania praniu pieniędzy (AML) czy finansowaniu terroryzmu (CFT). Dzięki algorytmom uczenia maszynowego instytucje finansowe mogą analizować miliardy transakcji i generować alerty na podstawie nieoczywistych wzorców zachowań, co ogranicza zarówno liczbę fałszywych alarmów, jak i ryzyko niewykrycia prawdziwego zagrożenia,

2. optymalizacja alertów i ustalanie progów – AI umożliwia dostrajanie progów alertów (threshold fine-tuning) w zależności od zmieniających się warunków rynkowych, co zmniejsza przeciążenie analityków dużą liczbą alertów do ręcznej weryfikacji,

3. zautomatyzowana analiza sieci powiązań – systemy AI mogą analizować powiązania pomiędzy klientami, kontrahentami i innymi uczestnikami transakcji, ujawniając złożone sieci i potencjalne działania przestępcze, np. w ramach tzw. network analysis,

4. zarządzanie ryzykiem klientów – AI wspiera ocenę ryzyka klienta w czasie rzeczywistym, wykorzystując dane o historii transakcji, źródłach dochodu, branży oraz geograficznych powiązaniach. Modele predykcyjne pozwalają przewidywać prawdopodobieństwo zaangażowania klienta w przestępstwa finansowe w horyzoncie 12–24 miesięcy,

5. weryfikacja beneficjenta i kontrola sankcyjna – w systemach płatności w czasie rzeczywistym AI wspomaga automatyczną kontrolę kontrahentów w kontekście sankcji międzynarodowych. Transakcje są weryfikowane pod kątem zgodności z listami sankcyjnymi jeszcze przed ich realizacją,

6. optymalizacja procesów dochodzeniowych – Natural Language Processing (NLP) automatyzuje tworzenie narracji dla śledczych i wspiera przygotowanie raportów z dochodzeń, co znacząco oszczędza czas i poprawia jakość raportowania,

7. adaptacja do regulacji i wymogów audytowych – Wraz z rosnącym nadzorem regulacyjnym, AI staje się narzędziem do spełniania wymogów

1 A. Saxena, S. Verma, J. Mahajan, *Generative AI in Banking Financial Services and Insurance*, Apress, 2024, s. 85–89.
2 C. Kerrigan, *Artificial Intelligence. Law and Regulation*, Elgar Publishing 2022, s. 327–328.
3 C. Kerrigan, dz. cyt., s. 267–268.
4 Saxena, S. Verma, J. Mahajan, dz. cyt., s. 86–87.
5 A. Gupta, D. N. Dwivedi, J. Shah, *Artificial Intelligence Applications in Banking and Financial Services*, Springer 2023, s. 6–9 oraz 65–114.

raportowych i audytowych, przy jednoczesnym zachowaniu efektywności operacyjnej.

Warto również podkreślić, że na początku roku 2025 P. Sobolewski zebrał rozwiązania AI stosowane w polskich bankach. Krajobraz krajowego wykorzystania AI wygląda następująco[6]:

1. **Nest Bank:**
   - **N! Asystent:** Jest to wirtualny asystent klienta dostępny w aplikacji mobilnej, który zdobył nagrody Best of Show na FinovateFall 2024 oraz TOP Innovation – AI na Banking Tech Awards 2024. Asystent wykorzystuje model GPT 4o, wzbogacony o wewnętrzne dokumenty banku (regulaminy, tabele opłat i prowizji), i jest połączony z usługami bankowości internetowej i mobilnej. Rozwiązanie umożliwia odpowiadanie na pytania dotyczące oferty, przygotowywanie analiz przychodów i wydatków, udzielanie informacji o historii transakcji dla klientów biznesowych. Klienci mogą komunikować się pisemnie i głosowo, zlecać przelewy, umawiać rozmowy z konsultantami, a także otrzymywać informacje finansowe na aktywnych wykresach. Ponadto w planach jest możliwość opłacania faktur poprzez zrobienie zdjęcia (na początku 2025 r.), rozszerzenie o funkcje przewalutowania, zastrzegania kart, składania wniosków o nowe karty oraz realizowania dyspozycji, np. zmiany danych.
   - **„Puchacz”**: wewnętrzny asystent AI oparty na modelu GPT, zaprojektowany w celu wsparcia pracowników banku, uruchomiony na początku 2023 r.
2. **VeloBank:**
   - **Chatbot wspierający doradców**: uruchomiony w październiku 2023 r. wspierał doradców obsługujących rządowy program Bezpieczny Kredyt 2 proc.. Bank planuje uruchomienie podobnych chatbotów wspierających sprzedawców innych produktów, np. pożyczek gotówkowych.
   - **Rekomendacje produktów ESG**: w listopadzie 2023 r. GenAI została wykorzystana do rekomendowania klientom produktów zgodnych z wartościami ESG na platformie sprzedażowej VeloMarket.
   - **VeloFotka**: uruchomiona we wrześniu 2024 r., umożliwia automatyczne wypełnianie wniosku o kredyt konsumpcyjny do 6,5 tys. zł poprzez zrobienie zdjęcia etykiety cenowej produktu. Wykorzystuje Chat-GPT 4 Turbo do analizy zdjęcia. Od wdrożenia bank udzielił finansowania na ponad 5 mln zł w ten sposób.
   - **Service bot dla pożyczki gotówkowej**: obecnie w testach, ma prowadzić klienta przez proces wnioskowania i być bardziej interaktywny niż dotychczasowe chatboty dla pracowników.
3. **PKO BP:**
   - bank intensywnie wykorzystuje klasyczne modele uczenia maszynowego (machine learning) na masową skalę, m.in. w:
     - Monitoringu sankcji.
     - Wykrywaniu oszustw.
     - Przeciwdziałaniu praniu pieniędzy.
     - Modelach wyceny (np. kart kredytowych).
     - Rozpoznawaniu intencji klientów.
     - Wydobywaniu danych i obsłudze klientów (np. realizacja przelewów, blokada kart, sprawdzanie zainteresowania produktem, analiza wydatków).
4. **BNP Paribas:**
   - **GENiusz**: wewnętrzny chatbot dla pracowników banku.
   - **CopAllot**: asystent kodowania dla programistów.
   - **AIDA**: asystent umożliwiający zadawanie pytań do baz danych języka programowania SQL przez osoby nieznające tego języka.
5. **Alior Bank:**
   - **Aliorpedia**: wewnętrzna baza wiedzy dla pracowników. Narzędzie to działa w formie chatu i umożliwia zadawanie pytań dotyczących wewnętrznych procedur i regulacji bankowych, udzielając odpowiedzi w języku naturalnym. Pełnoskalowe wdrożenie Aliorpedii planowane jest na 2025 rok.
6. **Bank Millenium:**
   - bank udostępnił pracownikom platformę opartą o GenAI, która wspiera ich w analizie dokumentów, obsłudze reklamacji i transkrypcji rozmów.

6 P. Sobolewski, *Kto, jak i dlaczego używa genAI w polskiej bankowości*, XYZ.pl, 2025.

7. ING:

- bank prowadzi prace pilotażowe, pod kątem:
  - Specjalistycznych chatbotów wspierających procesy w banku.
  - Botów dla pracowników bazujących na wiedzy o produktach i procesach w banku.

8. BOŚ (Bank Ochrony Środowiska)

- bank udostępnił czat produktowy dla doradców na potrzeby wewnętrzne.

9. Credit Agricole

- bank testuje narzędzie do upraszczania tekstów, a także planuje uruchomić w 2025 roku wirtualnego asystenta, który będzie przygotowywał dla doradcy kluczowe informacje o kliencie przed spotkaniem, tworzył notatki po spotkaniu na podstawie nagrania doradcy i dobierał produkty/usługi dla klienta. Innym planowanym rozwiązaniem jest czat dla pracowników wspierający obsługę klientów na podstawie wewnętrznych instrukcji i procedur.

10. Bank Pocztowy

- bank wdrożył rozwiązania wspierające pracowników w **procesach wewnętrznych**, np. w przygotowywaniu i opiniowaniu dokumentacji pod kątem zgodności z regulacjami wewnętrznymi i spójności merytorycznej, a także w weryfikacji dokumentów prawnych i audytowych.

Podsumowując, znaczenie AI w sektorze płatniczym rośnie, nie tylko na arenie międzynawowej, ale również w Polsce. Umożliwia ono instytucjom finansowym skuteczniejsze zarządzanie ryzykiem, szybsze wykrywanie zagrożeń, a także lepsze dostosowanie usług do indywidualnych potrzeb klientów. W miarę rozwoju technologii, rola sztucznej inteligencji w płatnościach będzie coraz bardziej fundamentalna dla zapewnienia bezpieczeństwa, dostępności i innowacyjności usług finansowych[7].

1.3. Aktualny stan regulacji – brak jednoznacznej definicji AI w prawie

Pomimo wejścia w życie aktu o sztucznej inteligencji (AI Act), obowiązującego od 1 sierpnia 2024 r., pojęcie „systemu AI" pozostaje nieostre i wieloznaczne, co stwarza liczne wyzwania regulacyjne[8].

Choć art. 3 ust. 1 AI Act zawiera formalną definicję systemu AI, wskazując, że jest to „system oparty na maszynie, zaprojektowany do działania z różnym poziomem autonomii, który może wykazywać zdolność adaptacyjną po wdrożeniu i który, w celu realizacji celów jawnych lub ukrytych, wnioskuje na podstawie danych wejściowych, jak wygenerować dane wyjściowe mogące wpływać na środowisko fizyczne lub wirtualne", definicja ta jest szeroka, ogólna i wielowarstwowa[9].

Komisja Europejska, zgodnie z art. 96 ust. 1 lit. f AI Act, opracowała szczegółowe wytyczne do tej definicji, wskazując na siedem podstawowych elementów (system oparty na maszynie, autonomia, adaptacyjność, celowość, proces wnioskowania, rodzaj danych wyjściowych i zdolność wpływania na środowisko), zaznaczając przy tym, że lista nie ma charakteru zamkniętego i każdy przypadek należy analizować indywidualnie[10].

W ocenie European Law Institute (ELI) taka konstrukcja prawna budzi poważne zastrzeżenia z punktu widzenia pewności prawa. Eksperci ELI wskazują, że obecna definicja systemu AI ma charakter ogólnikowy i nieprzejrzysty, co prowadzi do problemów interpretacyjnych w praktyce regulacyjnej oraz ryzyka nadmiernego lub niezamierzonego objęcia regulacją innych niż AI systemów algorytmicznych[11].

ELI proponuje tzw. „trójczynnikowe podejście" (Three-Factor Approach), które pozwala precyzyjniej odróżnić systemy AI od innych rozwiązań cyfrowych. Podejście to opiera się na: (1) stopniu wykorzystania danych lub wiedzy eksperckiej przy projektowaniu systemu; (2) zdolności systemu do generowania no-

7 GAO, *Artificial Intelligence: Use and Oversight in Financial Services*, GAO-25–107197, 2025, s. 8–10, A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 87–90.

8 Komisja Europejska, *Commission Guidelines on the Definition of an Artificial Intelligence System Established by Regulation (EU) 2024/1689 (AI Act), C (2025) 924 final*, Bruksela 2025 r., s. 1–2.

9 Tamże, s. 2.

10 Tamże, s. 2–3.

11 Krytyczna analiza nieprecyzyjności definicji prawnej oraz propozycja „Three-Factor Approach" (dane, tworzenie wiedzy, nieokreśloność wyjścia). *European Law Institute, Response on the Definition of an AI System: Commission Guidelines on the Application of the Definition of an AI System and the Prohibited AI Practices Established in the AI Act*, Wiedeń 2024 r., s. 19–21.

wej wiedzy podczas działania; (3) stopniu formalnej nieokreśloności generowanych wyników[12].

Trudności interpretacyjne związane z definicją systemu AI przekładają się także na praktykę oceny zgodności oraz wyznaczania obowiązków producentów, użytkowników i innych uczestników cyklu życia systemów AI. Nawet szczegółowe wytyczne Komisji nie pozwalają jednoznacznie rozstrzygnąć, które rozwiązania technologiczne rzeczywiście mieszczą się w zakresie zastosowania AI Act, a które nie[13].

Na ten problem zwraca również uwagę niemieckie Federalne Biuro ds. Bezpieczeństwa Informacji w Niemczech, które w katalogu testowym dla systemów AI w sektorze finansowym wskazuje, że brak jednoznaczności definicyjnej komplikuje wdrażanie wymogów oceny ryzyka i zgodności z przepisami AI Act, zwłaszcza w kontekście dynamicznie rozwijających się technik uczenia maszynowego i systemów generatywny[14].

Wytyczne etyczne Komisji Europejskiej z 2019 r. również zauważały brak szeroko uzgodnionej, jednoznacznej definicji systemów AI, podkreślając zróżnicowanie podejść akademickich, inżynieryjnych i prawnych, oraz wynikającą z tego potrzebę ostrożności regulacyjnej i etycznej w projektowaniu i stosowaniu tych systemów[15].

Podsumowując, brak jednoznacznej, technicznie i prawnie spójnej definicji AI w AI Act oraz w aktach wykonawczych prowadzi do ryzyka regulacyjnego, niepewności dla podmiotów gospodarczych oraz zagrożenia dla skuteczności egzekwowania przepisów w praktyce. Problem ten pozostaje aktualny i wymaga dalszej pracy interpretacyjnej, zarówno ze strony instytucji unijnych, jak i środowiska naukowego i praktyków prawa[16].

1.3.1. Porównanie definicji systemu AI z definicją OECD oraz wpływ normy ISO/IEC 42001

Warto zauważyć, że definicja systemu sztucznej inteligencji przyjęta w art. 3 ust. 1 AI Act została w dużym stopniu zaczerpnięta z definicji zaproponowanej przez OECD w 2023 r. Zgodnie z dokumentem OECD, system AI to: „a machine-based system that, for explicit or implicit objectives, infers from the input it receives how to generate outputs that can influence physical or virtual environments". Kluczowe elementy tej definicji — takie jak wnioskowanie, autonomia, celowość działania oraz wpływ na środowisko — zostały bezpośrednio przejęte przez prawodawcę unijnego, co podkreśla wysoki poziom internacjonalizacji pojęcia systemu AI w nowym porządku prawnym[17].

Jednocześnie warto podkreślić, że w odróżnieniu od AI Act, który ogranicza się do definicji prawnej, dokument OECD proponuje także podejście funkcjonalne i procesowe — traktując system AI jako technologię przechodzącą przez fazy budowy i użycia, z naciskiem na rolę „adaptive learning" i zmienności zachowań systemu w czasie. OECD uwzględnia również znaczenie odpowiedzialności i kontekstu implementacji, co staje się punktem wyjścia dla analizy ryzyka i klasyfikacji zastosowań AI[18].

W tym kontekście szczególną rolę pełni norma ISO/IEC 42001:2023 — pierwsza międzynarodowa norma systemu zarządzania AI (AI Management System – AIMS). ISO/IEC 42001 dostarcza ram organizacyjnych i operacyjnych, w których definicja systemu AI przestaje być wyłącznie techniczna, a staje się powiązana z procesami oceny ryzyka, etyki i zgodności. Norma nie tworzy własnej definicji AI, lecz opiera się na rozumieniu systemu AI jako zbioru komponentów (technicznych, procesowych, decyzyjnych), które mogą mieć wpływ na osoby, organizacje lub środowisko. Szczególny nacisk kładzie na identyfikowalność, zdolność do wyjaśniania, bezpieczeństwo funkcjonalne i zgodność z ramami regulacyjnymi — w tym z AI Act[19].

Co istotne, ISO/IEC 42001 uwzględnia różne poziomy zaawansowania systemów — od statycznych reguł decyzyjnych po uczenie maszynowe z dynamiczną adaptacją — i definiuje obowiązki organizacyjne

12 Tamże, 19–22.
13 Komisja Europejska, *Commission Guidelines on the Definition…*, dz. cyt., s.4–5.
14 Federal Office for Information Security, *Test Criteria Catalogue for AI Systems in Finance*, Bonn 2025, s. 2–7.
15 Grupa ekspertów wysokiego szczebla ds. SI, *Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji*, Komisja Europejska 2019, s. 48.
16 European Law Institute, dz. cyt., s. 26.
17 Tamże, s. 8.
18 Komisja Europejska odwołuje się bezpośrednio do dokumentu OECD jako podstawy definicyjnej oraz procesowej perspektywy systemów AI. Komisja Europejska, *Commission Guidelines on the Definition…*, dz. cyt., s. 2.
19 S. Musch, M. Borrelli, C. Kerrigan, J. Bennison, *EU AI Act ISO/IEC 42001: 2023. A Guide to Implementation*, AI & Partners 2025, s. 5–6.

związane z projektowaniem, wdrażaniem i monitorowaniem takich systemów. To ujęcie umożliwia lepsze praktyczne rozróżnienie między systemami, które powinny być objęte regulacją AI Act, a tymi które mogą być z niej wyłączone, co pozostaje niejasne przy zastosowaniu samej definicji prawnej[20].

Zatem w zestawieniu z podejściem OECD oraz strukturą ISO/IEC 42001, definicja z AI Act jawi się jako punkt wyjścia, lecz nie jako zamknięta klasyfikacja. Jej użyteczność praktyczna wymaga uzupełnienia o wymiar organizacyjny (ISO) oraz funkcjonalny (OECD), zwłaszcza w kontekście oceny zgodności, certyfikacji i audytu ex ante systemów wysokiego ryzyka.

1.3.2. Agenci AI i ryzyko systemowe

Definicja systemu AI przyjęta w AI Act wydaje się być szczególnie trudna do zastosowania w odniesieniu do tzw. agentów AI, czyli systemów zdolnych do autonomicznego realizowania złożonych celów w zmiennym środowisku. Raport „Ahead of the Curve: Governing AI Agents under the EU AI Act" przygotowany przez The Future Society wskazuje, że agenci AI łączą cechy systemów ogólnego przeznaczenia (GPAI) z funkcjonalną autonomią i adaptacyjnością, co znacząco komplikuje ich klasyfikację zgodnie z aktualnym brzmieniem art. 3 AI Act[21].

Agenci AI — rozumiani jako systemy zdolne do samodzielnego podejmowania decyzji w środowiskach wirtualnych i fizycznych — budzą obawy nie tylko z punktu widzenia klasyfikacji prawnej, lecz przede wszystkim ryzyka systemowego. Raport wskazuje, że w przypadku agentów opartych na modelach GPAI z ryzykiem systemowym (GPAISR), trudno jest z góry ocenić, czy konkretne wdrożenie będzie spełniać przesłanki „systemu wysokiego ryzyka" z rozdziału III AI Act, czy też nie. Wszystko zależy od kontekstu użycia[22].

Jednocześnie w raporcie podniesiono, że z punktu widzenia skutków społecznych i prawnych, nawet jeśli agenci nie zostaną formalnie uznani za systemy wysokiego ryzyka, mogą i tak generować złożone, rozproszone ryzyka, charakterystyczne dla zjawiska tzw. „wielu rąk" (many hands problem). Ryzyko systemowe pojawia się m.in. w kontekście koluzji wieloagentowej, samonapędzających się błędów operacyjnych, czy trudności w atrybucji odpowiedzialności w całym łańcuchu dostaw[23].

Future Society proponuje więc, aby do takich systemów stosować rozszerzone podejście regulacyjne oparte na czterech filarach: ocenie ryzyka, narzędziach przejrzystości, kontrolach technicznych oraz nadzorze człowieka. Co istotne, rozkłada ono odpowiedzialność regulacyjną pomiędzy dostawców modeli GPAI, dostawców systemów końcowych oraz użytkowników wdrażających — podkreślając potrzebę kontekstowej interpretacji definicji AI systemu i odejścia od czysto formalnych klasyfikacji[24].

W świetle powyższego, agenci AI stanowią szczególne wyzwanie dla obecnej definicji systemu AI — ze względu na swoją złożoność, elastyczność i trudną do przewidzenia ścieżkę interakcji ze środowiskiem. Przypadek ten unaocznia potrzebę dalszej precyzji w stosowaniu definicji AI systemów, z uwzględnieniem ich wpływu, a nie tylko deklarowanej funkcji technicznej[25].

20 Tamże, s. 10–12.
21 A. Oueslati, R. States-Polet, *Ahead od the Curve: Governing AI Agents Under the EU AI Act*, The Future Society 2025, s. 1–2.
22 Tamże, s. 22–23.
23 Tamże, s. 18 oraz 28.
24 Tamże, s. 30–32.
25 Tamże, s. 5 oraz 33.

2. Automatyzacja procesów płatniczych dzięki AI

Dotychczasowe wdrożenia sztucznej inteligencji w sektorze finansowym koncentrowały się głównie na efektywności operacyjnej, wykrywaniu anomalii, automatyzacji i zwiększaniu skali obsługi. Jednak z perspektywy interesariuszy — klientów indywidualnych i regulatorów — coraz większe znaczenie zyskują społeczne skutki tych technologii. Raporty, tak krajowe, jak i zagraniczne, pokazują, że choć główne motywacje wdrażania AI w sektorze finansowym to automatyzacja, wykrywanie nadużyć i optymalizacja kosztów, coraz większą wagę przywiązuje się do społecznych konsekwencji tych działań — takich jak zaufanie, inkluzywność, przejrzystość i ochrona prywatności klientów[26].

Wyniki badań przeprowadzonych w Polsce pokazują, że większość klientów nie ma świadomości, że już korzysta z usług opartych na AI, a jednocześnie wyraża silną potrzebę zachowania kontaktu z człowiekiem, zwłaszcza w sytuacjach problemowych lub związanych z pieniędzmi. Dla wielu użytkowników kluczowe są[27]:

- transparentność działania systemów AI,
- łatwość rezygnacji z interakcji z AI na rzecz człowieka,
- wyjaśnialność podejmowanych decyzji,
- oraz poszanowanie prywatności i danych osobowych.

Systemy AI, szczególnie te używane do scoringu kredytowego czy detekcji fraudów, mogą nieświadomie powielać uprzedzenia i prowadzić do dyskryminacji grup społecznie wrażliwych. Dotyczy to np. osób starszych, migrantów, kobiet czy osób o niższym statusie cyfrowym. W tym kontekście istotne jest wprowadzenie mechanizmów[28]:

- audytów etycznych,
- walidacji danych treningowych,
- oraz monitorowania tzw. fairness AI na etapie wdrożenia i eksploatacji

Rozbudowa systemów opartych na AI może paradoksalnie zwiększać wykluczenie cyfrowe, jeśli nie zostaną zapewnione alternatywne kanały obsługi oraz odpowiednie wsparcie dla klientów mniej biegłych technologicznie. Według raportu UNDP (2025), adaptacja AI powinna być równoważona przez inwestycje w kompetencje cyfrowe obywateli i budowanie społecznej odporności na zmianę[29].

Skuteczne wdrażanie AI i automatyzacja w wymaga nie tylko zaawansowanej technologii, ale również zrozumienia społecznego kontekstu jej użycia. Równoważenie perspektywy technologicznej z podejściem zorientowanym na człowieka (human-centric AI) jest warunkiem społecznej akceptowalności, zgodności regulacyjnej oraz długofalowej skuteczności rozwiązań AI[30].

26 K. Sekścińska, *AI w finansach i bankowości oczami polskiego konsumenta*, SYGN. PAB/WIB 7/2025, s. 6–7; UNDP, *Human Development Report 2025 – A Matter of Choice: People and Possibilities in the Age of AI*, United Nations Development Programme 2025, s. v– vi.

27 Badanie przeprowadzone na ogólnopolskiej próbie pokazuje, że 52% respondentów nie wie, że korzystało z AI, a 93% preferuje kontakt z człowiekiem w sprawach finansowych. Klienci wskazują, że kluczowe dla akceptacji AI są: przejrzystość działania systemów, możliwość szybkiego przełączenia na obsługę ludzką, wyjaśnialność decyzji oraz pełne poszanowanie prywatności danych. K. Sekścińska, dz. cyt, s. 26–28.

28 Zidentyfikowane w badaniach w USA i UE ryzyka obejmują powielanie uprzedzeń i dyskryminację w decyzjach kredytowych i ocenie ryzyka, zwłaszcza wobec grup wrażliwych. W rekomendacjach pojawia się potrzeba stosowania audytów etycznych, oceny danych treningowych oraz systematycznego monitorowania zgodności modeli z zasadami równości i niedyskryminacji. GAO, *Artificial Intelligence: Use and Oversight in Financial Services*, GAO-25–107197, 2025, s. 8–9; CEPS, *Analysis of EU AI Office Stakeholder Consultations: defining AI systems and prohibited applications*, CEPS, 2025, s. 32–36.

29 Na przykład w raporcie UNDP podkreślono, że szybka adaptacja AI bez równoległych inwestycji w rozwój kompetencji cyfrowych, infrastruktury wspierającej i inkluzyjnych polityk może pogłębiać istniejące wykluczenia. Kluczowe znaczenie ma budowanie społecznej odporności na zmianę technologiczną i zapewnianie alternatywnych ścieżek dostępu do usług publicznych i finansowych. UNDP, dz. cyt., s. 45–47.

30 Grupa ekspertów wysokiego szczebla ds. SI, Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji, Komisja Europejska 2019, s. 4–6; UNDP, dz. cyt., s. 14–15.

2.1. Mechanizmy automatyzacji w płatnościach online oraz w płatnościach mobilnych

Transformacja cyfrowa sektora finansowego w ostatnich latach przyspieszyła dzięki wykorzystaniu sztucznej inteligencji (AI). Jednym z kluczowych obszarów, gdzie zastosowanie tej technologii przynosi wymierne korzyści, jest automatyzacja procesów płatniczych i przetwarzania transakcji. Tradycyjne mechanizmy przetwarzania, oparte na regułach i z góry zdefiniowanych procedurach, stają się niewystarczające w obliczu dynamicznych zagrożeń, rosnących wolumenów danych oraz oczekiwań klientów. AI umożliwia wdrożenie adaptacyjnych, skalowalnych i bezpiecznych rozwiązań, które znacząco poprawiają efektywność i redukują koszty operacyjne[31].

Zastosowanie sztucznej inteligencji w procesach płatniczych polega przede wszystkim na analizie danych transakcyjnych w czasie rzeczywistym, identyfikacji wzorców zachowań klientów, wykrywaniu anomalii oraz wspomaganiu podejmowania decyzji. W praktyce oznacza to m.in. automatyczną klasyfikację transakcji, ocenę ich zgodności z regulacjami wewnętrznymi i zewnętrznymi oraz przewidywanie ewentualnych nieprawidłowości[32].

AI, a szczególnie techniki uczenia maszynowego (machine learning, ML), umożliwiają przetwarzanie milionów rekordów w czasie rzeczywistym, przy jednoczesnym uczeniu się i doskonaleniu algorytmów na podstawie napływających danych. Banki komercyjne stosują tego typu systemy w obszarach takich jak przetwarzanie płatności kartowych, przelewów międzybankowych (np. SWIFT), rozliczeń masowych czy oceny wiarygodności płatnika[33].

Jednym z wiodących przykładów wdrożenia AI do automatyzacji transakcji jest Bank of America, który korzysta z rozwiązań opartych na sztucznej inteligencji do przetwarzania wniosków kredytowych i analizy historii płatności klientów. Podobnie, JP Morgan Chase wdrożył system COIN (Contract Intelligence), który analizuje dokumentację prawną powiązaną z transakcjami finansowymi w ułamku czasu potrzebnego pracownikowi[34].

Również HSBC, w kontekście przeciwdziałania praniu pieniędzy (AML), wykorzystuje AI do weryfikacji transakcji pod kątem nieprawidłowości i podejrzanych schematów przepływu środków. Dzięki temu możliwe jest automatyczne wyselekcjonowanie przypadków wymagających interwencji analityka, co zmniejsza nakład pracy ludzkiej nawet o 60%[35].

Automatyzacja przynosi liczne korzyści, takie jak[36]:

- redukcja błędów ludzkich, które często występują przy ręcznym przetwarzaniu;
- znaczne przyspieszenie procesów, co pozwala realizować przelewy w czasie rzeczywistym;
- obniżenie kosztów operacyjnych dzięki zmniejszeniu zapotrzebowania na pracę manualną;
- zwiększenie zgodności regulacyjnej, poprzez automatyczne uwzględnianie aktualnych przepisów prawa i wewnętrznych polityk banku.

Co istotne, systemy te uczą się na podstawie historycznych danych transakcyjnych, co pozwala im lepiej przewidywać ryzyka i wychwytywać nieoczywiste schematy oszustw[37]. W przypadku generatywnej AI zdolność ta zyskuje nowy wymiar. Wysoka zdolność adaptacyjna GenAI – czyli bieżące „uczenie się" na nowych danych – podnosi skuteczność detekcji anomalii, ale także zwiększa potrzebę ciągłego mo-

31 A. Saxena, S. Verma, J. Mahajan, *Generative AI in Banking Financial Services and Insurance*, Apress, 2024, s. 59–66.
32 Tamże, s. 90–91.
33 Tamże, s. 91–94.
34 Tamże, s. 92–94.
35 Tamże, s. 113.
36 Tamże, 94–96.
37 Tamże, s. 96. Ponadto dokument „Automating Deception AI's Evolving Role in Romance Fraud" wydany przez The Alan Turing Institute opisuje, że modele językowe (LLM) są wykorzystywane do analizy narracji ofiar oszustw miłosnych poprzez przetwarzanie danych historycznych, takich jak zgłoszenia i wiadomości. Wykorzystano m.in. analizę częstości słów (TF-IDF) oraz modelowanie tematów (LDA), co pozwoliło na identyfikację powtarzających się wzorców językowych i typowych ról oszustów. Modele te mogą także przewidywać nowe warianty oszustw i wykrywać schematy na podstawie wcześniejszych danych. Podobnie, inny dokument wydany przez The Alan Turing Institute „AI and Serious Crime" odnosi się do badania, które wykorzystuje dane historyczne (np. narracje ofiar) i modele językowe do generowania oraz oceny wiadomości oszukańczych w celu lepszego zrozumienia i przewidywania oszustw z użyciem AI. Zob. S. Moseley, Automating Deception AI's Evolving Role in Romance Fraud, The Alan Turing Institute 2025, s. 3 i n.; J. Burton, A. Janjeva, S. Moseley, A. Moseley, AI and Serious Crime, The Alan Turing Institute 2025, s. 4 i n.

nitorowania, ponownej walidacji i ścisłego zarządzania zakresem samouczenia systemów[38].

Pomimo licznych zalet, wdrażanie AI w automatyzacji transakcji niesie również wyzwania. Kluczowe z nich to[39]:

- zapewnienie przejrzystości i wyjaśnialności działania algorytmów (tzw. explainable AI);
- ryzyko błędnych klasyfikacji (false positives), które mogą prowadzić do zablokowania prawidłowych płatności;
- konieczność ciągłego dostosowywania modeli do zmieniających się warunków rynkowych i regulacyjnych;
- zapewnienie zgodności z przepisami o ochronie danych osobowych (np. RODO/GDPR)
- zapewnienie zgodności z przepisami dotyczącymi cyberbezpieczeństwa w tym NIS 2 czy DORA.

Dodatkowo, wysokie koszty początkowe implementacji zaawansowanych rozwiązań AI mogą stanowić barierę dla mniejszych instytucji finansowych[40].

Zgodnie z prognozami PwC przytoczonymi w raporcie „A Playbook for Crafting AI Strategy", wdrożenie AI w sektorze finansowym ma potencjał, by zwiększyć globalne PKB o 15,7 biliona dolarów do 2030 roku, z czego znaczna część pochodzić będzie właśnie z automatyzacji usług płatniczych i procesów transakcyjnych. Jednocześnie wzrasta znaczenie AI w kontekście zgodności regulacyjnej (RegTech), gdzie algorytmy pełnią rolę nie tylko strażnika zgodności, lecz także doradcy biznesowego[41].

Automatyzacja przetwarzania transakcji za pomocą AI to obecnie jeden z najważniejszych kierunków transformacji sektora bankowego. Przynosi ona znaczne usprawnienia operacyjne, wzmacnia bezpieczeństwo i pozwala instytucjom finansowym skuteczniej konkurować na rynku. Warunkiem skutecznego wykorzystania AI pozostaje jednak odpowiednia jakość danych, właściwa architektura systemowa oraz dbałość o transparentność i zgodność z przepisami prawa[42] oraz cyberbezpieczeństwo.

2.2. Wykorzystanie AI w analizie transakcji i zarządzaniu ryzykiem

W obliczu rosnącej liczby transakcji finansowych i coraz bardziej złożonych zagrożeń instytucje finansowe muszą nie tylko przetwarzać dane szybko i precyzyjnie, ale także zarządzać wielowymiarowym ryzykiem: operacyjnym, kredytowym, reputacyjnym i regulacyjnym. W tym kontekście sztuczna inteligencja (AI) odgrywa kluczową rolę – umożliwia identyfikację nieprawidłowości, przewidywanie zagrożeń i wspiera decyzje zarządcze w sposób niedostępny dla tradycyjnych metod statystycznych[43].

Jednym z najbardziej zaawansowanych zastosowań AI w finansach jest **automatyczna analiza wzorców transakcyjnych** w czasie rzeczywistym. Dzięki technikom uczenia nienadzorowanego (unsupervised learning) i modelom behawioralnym możliwe jest wykrywanie odchyleń od standardowych zachowań klienta, co ma kluczowe znaczenie w walce z oszustwami finansowymi (fraud detection). Systemy te identyfikują sygnały ostrzegawcze dla fałszywych przelewów, kradzieży tożsamości czy nadużyć kartowych[44]. Wysoka zdolność adaptacyjna GenAI – czyli bieżące „uczenie się" na nowych danych – podnosi

38 Hong Kong Institute for Monetary and Financial Research, *Financial Services in the Era of Generative AI Facilitating Responsible Adoption*, HKIMR Applied Research Report No.1/2025, s. 10.
39 A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 98–99.
40 Tamże, s. 99. Ponadto poparcie to stwierdzenie znajdziemy w innych źródłach. Na przykład C. Keerigan pisze, że „Mniejsze firmy mogą być bardziej skłonne do przyjmowania rozwiązań oferowanych przez zewnętrznych dostawców. Dostawcy tego rodzaju usług zaczynają również oferować kompletne pakiety „sztucznej inteligencji jako usługi" (AI as a Service), które umożliwiają firmom korzystanie z rozwiązań opartych na AI bez konieczności ponoszenia dużych inwestycji we własną infrastrukturę.". C. Kerrigan, Artificial Intelligence. Law and Regulation, Elgar Publishing 2022, s. 28–29. Warte też zwrócić uwagę na kwestie kosztów ponownego trenowania: „Ponowne trenowanie modeli sztucznej inteligencji to kosztowny proces, dlatego instytucje finansowe często wykazują tendencję do opóźniania takich działań (tzw. inaction bias), co skutkuje dalszym stosowaniem błędnych metod.", A. T. da Fonseca, E. Vaz de Sequeira, L. B. Xavier, Liability for AI Driven Systems [w:] H. S. Antunes, P. M. Freitas, A. L. Oliveira, C. M. Pereira, E. V. de Sequeira, L. B. Xavier (red.), Multidisciplinary Perspectives on Artificial Intelligence and the Law, Vol. 58, Springer 2024, s. 300.
41 MIT Technology Review Insights, *A Playbook for Crafting AI Strategy*, 2024, s. 4.
42 A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 100.
43 Tamże, 124–125.
44 Tamże, 90–91.

skuteczność detekcji anomalii, ale także zwiększa potrzebę ciągłego monitorowania, ponownej walidacji i ścisłego zarządzania zakresem samouczenia systemów, aby zapobiec dryfowi modeli (model drift) i efektowi halucynacji decyzyjnych[45].

Modele detekcji anomalii są również wykorzystywane do przeciwdziałania praniu pieniędzy (AML), analizując złożone sieci powiązań i przepływów środków. Przykładowo, HSBC wdrożył systemy AI, które automatycznie identyfikują transakcje wysokiego ryzyka i klasyfikują je pod kątem obowiązku raportowania[46].

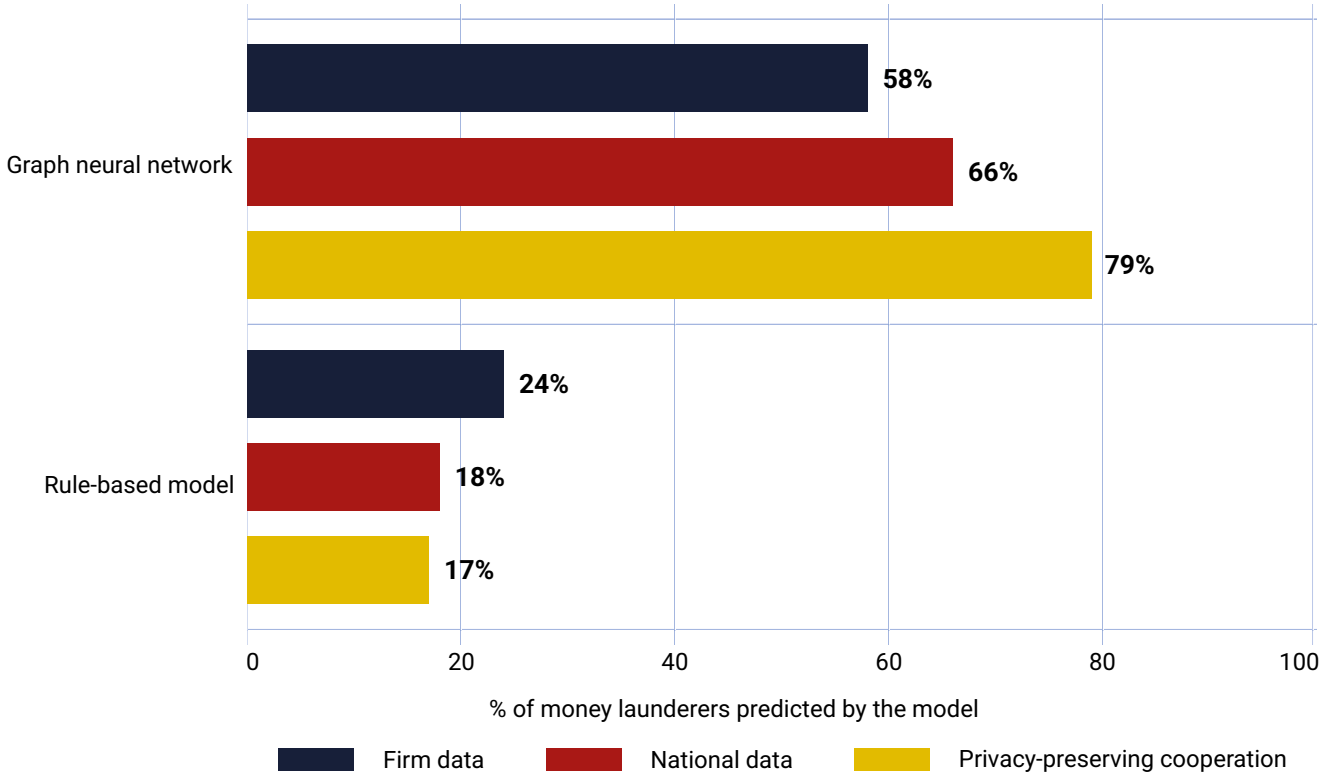
Na przykład Projekt Aurora realizowany przez BIS Innovation Hub wykorzystuje symulowane przypadki działań związanych z praniem pieniędzy, które są „wstrzykiwane" do danych płatniczych. Projekt porównuje skuteczność różnych narzędzi uczenia maszynowego z dominującym podejściem opartym na regułach, aby ocenić, w jakim stopniu poszczególne metody potrafią wykryć przypadki prania pieniędzy.

Porównanie przeprowadzono w trzech scenariuszach:

1. gdy dane transakcyjne są przetwarzane wyłącznie w obrębie pojedynczego banku (tj. w silosach);
2. przy wspólnym przetwarzaniu danych na poziomie krajowym;
3. w ramach współpracy transgranicznej z zastosowaniem metod ochrony prywatności, które nie ujawniają danych źródłowych organom w innych jurysdykcjach.

Wyniki pokazują, że modele oparte na uczeniu maszynowym przewyższają tradycyjne metody oparte na regułach, powszechnie stosowane w większości państw (zob rys. 1). Wspólne przetwarzanie danych na poziomie krajowym dodatkowo poprawia skuteczność. Najbardziej zaskakujące jest to, że metody uczenia maszynowego osiągają najwyższą skuteczność, gdy dane z różnych jurysdykcji są współdzielone w sposób zapewniający ochronę prywatności. Współpraca w zakresie danych znacząco zwiększa efektywność wykrywania w porównaniu z obecnymi metodami opartymi na regułach.

Rys. 1. Narzędzia sztucznej inteligencji w użyciu – dane transakcyjne widoczne na trzech różnych poziomach analizy: dane przedsiębiorstwa, dane krajowe oraz współpraca z zachowaniem prywatności.



Źródło: *BIS Innovation Hub, Project Aurora: the power of data, technology and collaboration to combat money laundering across institutions and borders, BIS 2023.*

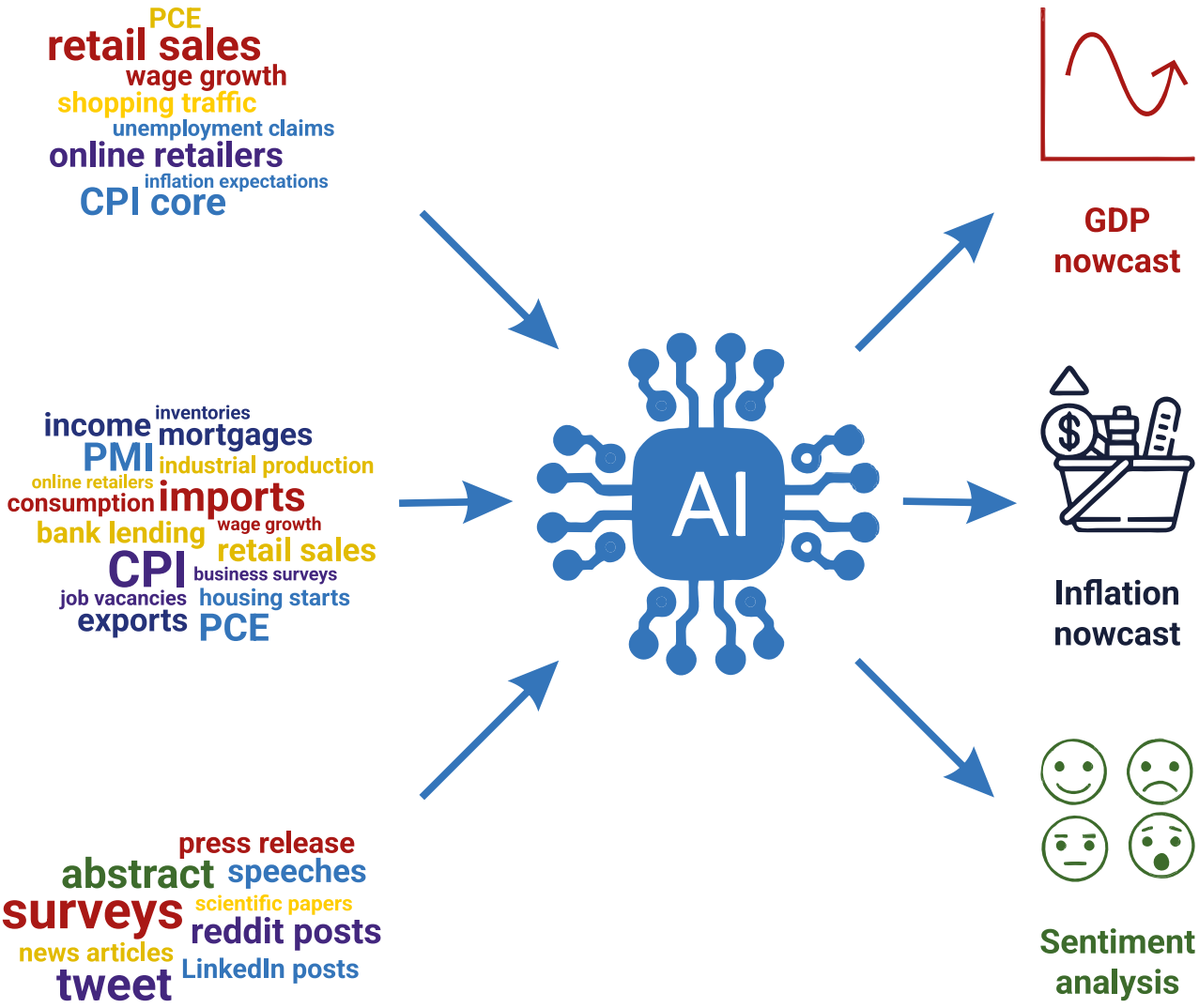
45 Swiss Banking, *Generative AI in Banking – A Comprehensive Overview What do banks need to consider before jumping into the rabbit hole?*, Expert report of the SBA 2025, s. 15.
46 Tamże, s. 113.

Oprócz analizy pojedynczych transakcji, AI wspiera **prognozowanie zdolności kredytowej klientów** oraz ryzyka niewypłacalności. Zamiast polegać wyłącznie na danych historycznych i wskaźnikach finansowych, modele AI integrują informacje z zachowań transakcyjnych, danych alternatywnych (np. rachunków za media), a nawet analizy sentymentu w kontaktach z klientem[47].

Tego typu rozwiązania są coraz częściej wykorzystywane przez banki detaliczne i fintechy, które potrzebują szybkiej, skalowalnej oceny ryzyka kredytowego. Zautomatyzowana segmentacja klientów i indywidualna ocena ryzyka poprawiają nie tylko jakość decyzji kredytowych, ale i zgodność z zasadą proporcjonalności w udzielaniu świadczeń finansowych[48].

Jednym z kluczowych zastosowań dużych modeli językowych (LLM) staje się obecnie też tzw. nowcasting – czyli bieżące prognozowanie aktywności gospodarczej lub inflacji. Tradycyjnie proces ten był ograniczany przez brak aktualnych danych oraz konieczność projektowania i trenowania modeli do bardzo wąsko zdefiniowanych zadań. Duże modele językowe mogą przełamać te ograniczenia dzięki swojej uniwersalności (zob. rys. 2). Jako tzw. zero-shot learners, potrafią generować prognozy bez potrzeby wcześniejszego dostrajania, co pozwala im „odkrywać igły w nieznanych stogach siana". Tak jak są trenowane do przewidywania kolejnego słowa w zdaniu, korzystając z ogromnych zbiorów danych tekstowych, tak samo mogą zostać wykorzystane do prognozowania kolejnych obserwacji numerycznych w szeregach czasowych[49].

Rys. 2. Modele typu „zero-shot" mogą zwiększyć możliwości modeli prognozowania.



Źródło: *H. S. Shin, Artificial Intelligence and the economy: implications for central banks, BIS 2024.*

47 Tamże, s. 126–127.
48 Tamże, s. 127.
49 H. S. Shin, *Artificial Intelligence and the economy: implications for central banks*, BIS 2024, s. 3–4.

Z punktu widzenia AI każdy typ danych – tekst, liczby, czy obrazy – to po prostu ciąg liczb. Dlatego też zdolności do rozpoznawania wzorców, które umożliwiają przewidywanie słów w języku naturalnym, mogą być z równym powodzeniem zastosowane do analizy i prognozowania zjawisk makroekonomicznych.

Wdrażanie AI w analizie transakcji nie może być oderwane od wymogów prawnych i etycznych. Zgodnie z AI Act i standardem ISO/IEC 42001, instytucje finansowe muszą wykazać, że ich systemy AI są transparentne, wyjaśnialne i zgodne z obowiązującymi przepisami[4]. Dotyczy to szczególnie systemów klasyfikowanych jako „high-risk", czyli m.in. wykorzystywanych w scoringu kredytowym czy analizie ryzyka klienta[50].

AI może automatyzować generowanie raportów zgodności, ułatwiając dokumentowanie przypadków podejrzeń o nadużycia i wspierając audyt ścieżek decyzyjnych. Wdrożenie takich systemów nie zwalnia jednak organizacji z obowiązku przeprowadzania oceny wpływu (impact assessment) i zapewnienia adekwatnych mechanizmów kontroli[51]. Skuteczna analiza transakcji za pomocą AI wymaga ścisłego powiązania ryzyka modeli z szerszym ładem nadzorczym nad jakością, integralnością i bezpieczeństwem danych – co obejmuje m.in. rozbudowane mechanizmy kontroli dostępu, filtrowanie treści wejściowych i nadzór nad źródłami danych[52].

Zastosowanie AI niesie również nowe ryzyka, związane z samymi modelami: nieprzejrzystość decyzji (black-box effect), nadmierna zależność od danych historycznych, a także ryzyko nieetycznych lub uprzedzonych decyzji. Zgodnie z ISO/IEC 42001 organizacje powinny regularnie przeglądać i testować modele AI pod kątem ich dokładności, odporności i uczciwości[53].

Ponadto, w kontekście generatywnej AI (GenAI), zarządzanie ryzykiem modeli staje się jeszcze bardziej złożone – wymaga uwzględnienia nie tylko dokładności predykcji, ale też przejrzystości danych uczących, kontroli nad cyklem życia modelu i mechanizmów awaryjnych (fallbacks) na wypadek błędów systemowych[54].

W praktyce zarządzanie ryzykiem modelu (model risk management) staje się integralną częścią funkcji compliance i audytu wewnętrznego. Wymaga to nowej kompetencji – łączenia wiedzy statystycznej z prawną i etyczną oceną działania systemu. W przeciwnym razie nawet technicznie skuteczne rozwiązanie może prowadzić do naruszenia przepisów lub szkody dla klienta[55].

AI znacząco zmienia podejście do zarządzania ryzykiem w sektorze finansowym – nie tylko automatyzując analizę transakcji, ale także przekształcając rolę audytu, compliance i oceny kredytowej. Jej skuteczność zależy jednak nie tylko od algorytmu, ale także od jakości danych, przejrzystości procesu i gotowości organizacji do wdrożenia zasad etycznego i odpowiedzialnego zarządzania[56].

2.3. AI w obsłudze klienta – chatboty i asystenci w bankowości

W dobie cyfryzacji usług finansowych coraz większą rolę odgrywają interfejsy konwersacyjne, oparte na chatbotach i wirtualnych asystentach. Wykorzystanie sztucznej inteligencji (AI), w tym przetwarzania języka naturalnego (NLP), pozwala instytucjom finansowym znacząco poprawić jakość obsługi klienta, usprawnić procesy płatnicze i zwiększyć skalowalność operacyjną. W szczególności chatboty i asystenci AI stają się istotnym elementem nowoczesnej architektury bankowości detalicznej i mobilnej[57].

Chatboty i wirtualni asystenci finansowi pełnią różnorodne funkcje – od prostych odpowiedzi na zapytania klientów, przez inicjowanie i zatwierdzanie transakcji, po zaawansowane doradztwo finansowe (niekiedy pomysły na nie są bardzo kreatywne, zob. rys.3). Ich zadania obejmują m.in.: informowanie o stanie konta i historii płatności, realizację przelewów i doładowań, przypomnienia o nadchodzących płatnościach, rekomendacje produktów i usług bankowych, weryfikację tożsamości i wsparcie KYC. Za-

50 Ministry of Infrastructure and Water Management, *AI Impact Assessment – The tool for a responsible AI project, Ver. 2.0*, 2024, s. 16–21; S. Musch, M. Borrelli, C. Kerrigan, J. Bennison, EU AI Act ISO/IEC 42001: 2023. *A Guide to Implementation*, AI & Partners 2025, s. 5–6.
51 Ministry of Infrastructure and Water Management, dz. cyt., s. 6–8.
52 Hong Kong Institute for Monetary and Financial Research, dz. cyt, s. 15.; Swiss Banking, dz. cyt, s. 14–15.
53 S. Musch, M. Borrelli, C. Kerrigan, J. Bennison, dz. cyt., s. 13–14.
54 Hong Kong Institute for Monetary and Financial Research, dz. cyt, s. 19 i n.; Swiss Banking, dz. cyt., s. 10–11.
55 Tamże,
56 Tamże, s. 14.
57 A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 107–108.

awansowani asystenci AI są w stanie prowadzić kontekstowe konwersacje, interpretować emocje klienta i dostosowywać ton komunikacji do sytuacji[58].

Jednym z najczęściej cytowanych wdrożeń jest asystent głosowy Erica, stworzony przez Bank of America. Innym przykładem jest Eno, asystent AI Capital One, który za pomocą SMS-ów i aplikacji mobilnej pomaga klientom monitorować wydatki. Klarna wprowadziła generatywnego asystenta AI opartego na ChatGPT-4, który już w pierwszym miesiącu obsłużył ponad 2/3 zapytań klientów[59].

Rys. 3. Nietypowy przykład zastosowania asystenta AI do zajmowania czasu oszustom stosującym vishing.

Źródło: https://youtu.be/RV_SdCfZ-0s

Zastosowanie chatbotów znacząco skraca czas reakcji i zwiększa dostępność usług – są one dostępne 24/7. Automatyzacja odpowiedzi pozwala ograniczyć obciążenie call center i zasobów ludzkich. Badania wykazują, że odpowiednio zaprojektowane asystenci AI mogą utrzymać poziom satysfakcji klienta na poziomie 80–90%[60].

Pomimo licznych korzyści, wdrożenie chatbotów i asystentów AI wiąże się z zagrożeniami: ograniczenia w rozumieniu kontekstu, ryzyko błędnej interpretacji, potrzeba zabezpieczenia komunikacji i zapewnienia zgodności z regulacjami (np. RODO, DORA) czy cyberbezpieczeństwo agentów. Standard ISO/IEC 42001 podkreśla znaczenie nadzoru i przejrzystości działania systemów AI, wskazuje się także na inne elementy przykładowo ocena Ministerstwa Infrastruktury i Gospodarki Wodnej w Niderlandach wskazuje na konieczność uwzględnienia wpływu takich systemów na prawa użytkowników[61].

Kierunek rozwoju chatbotów zmierza ku modelom multimodalnym, które oprócz tekstu interpretują głos i dane biometryczne. W kontekście globalnym, jak podkreśla to raport UNDP, technologia AI może działać jako narzędzie wspierające człowieka – pod warunkiem, że wdrażana jest odpowiedzialnie i etycznie[62].

58 Tamże, 108–111.
59 Tamże, 111–112.
60 MIT Technology Review Insights, dz. cyt., s. 10–12.
61 Ministry of Infrastructure and Water Management, dz. cyt, s. 16–21; S. Musch, M. Borrelli, C. Kerrigan, J. Bennison, dz. cyt., s. 5–6.
62 UNDP, dz. cyt., s. v– vi.

2.3.1. AI agenci w automatyzacji procesów finansowych

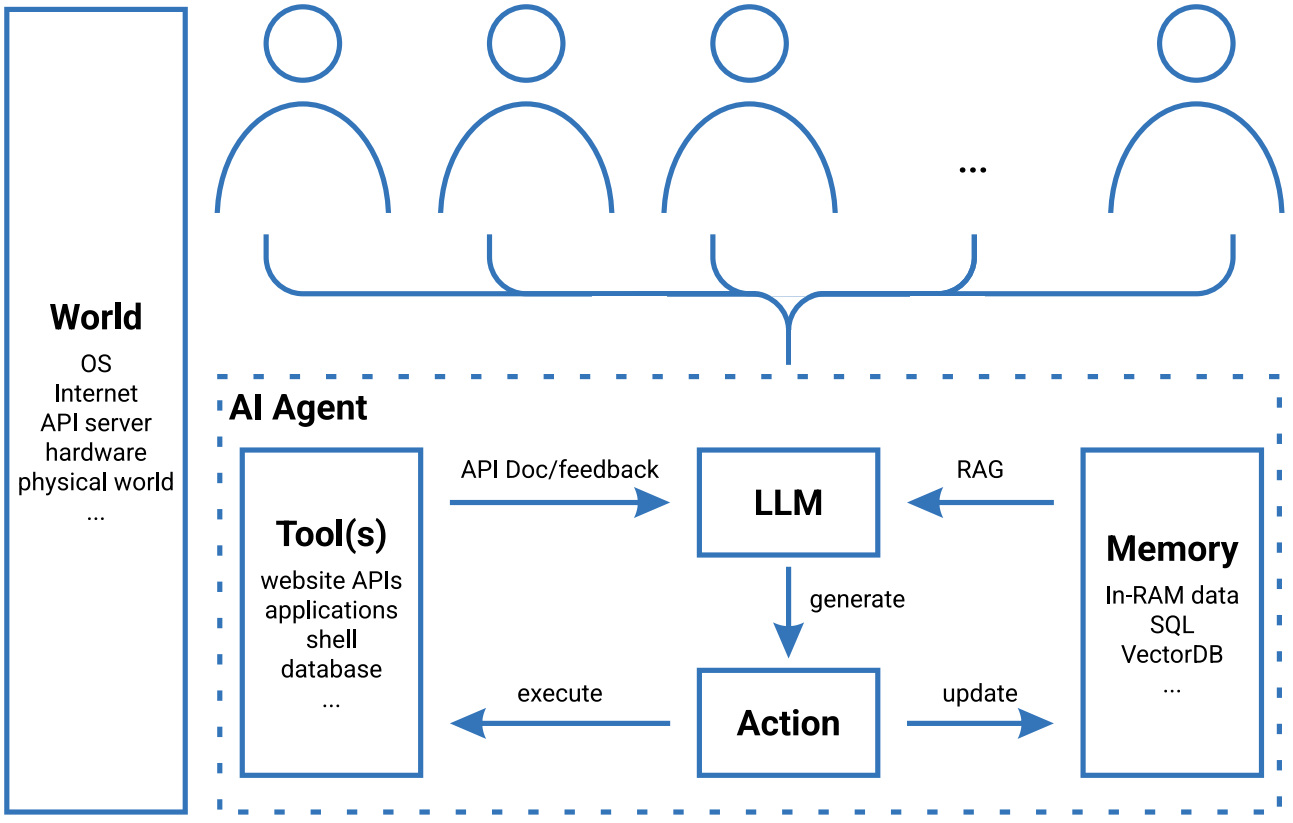
Wraz ze wzrostem wolumenu danych oraz złożoności procesów decyzyjnych, instytucje finansowe coraz częściej sięgają po sztuczną inteligencję nie tylko do analizy ryzyka czy detekcji oszustw, ale również do automatyzacji działań operacyjnych i interakcji z klientem. Obok klasycznych rozwiązań opartych na machine learningu i systemach regułowych, coraz większe znaczenie zyskują tzw. asystenci AI i agenci AI – systemy zdolne nie tylko do prowadzenia rozmowy czy generowania rekomendacji, ale również do samodzielnego planowania, interakcji z wieloma systemami i podejmowania decyzji[63].

Ta nowa generacja rozwiązań, bazująca na dużych modelach językowych (LLM) oraz tzw. architekturach agentowych, wprowadza istotną zmianę jakościową w podejściu do automatyzacji procesów w finansach. W kolejnej części przyjrzymy się bliżej różnicom między chatbotami, asystentami AI a agentami, oraz potencjalnym zastosowaniom tych ostatnich w sektorze finansowym[64].

Czym są AI agenci? AI agenci to systemy oparte na modelach językowych, które[65] (zob. rys.4):

- samodzielnie planują działania (np. wykonują wiele kroków w określonym celu),
- komunikują się z innymi systemami lub agentami (np. API, bazami danych),
- mogą wykonywać interakcje z klientami, generować dokumenty lub analizować dane.

Rys. 4. Przegląd agenta AI opartego na LLM.



Źródło: *Y. He, E. Wang, Y. Rong, H. Chen, Z. Cheng, Security of AI Agents*, https://arxiv.org/html/2406.08689v2#S1.

63 J. Kraprayoon, Z. Williams, R. Fayyaz, *AI Agents Governance: A Field Guide*, IAPS 2025, s. 1–3.

64 Źródła wskazują, że rozwój dużych modeli językowych (LLM) oraz systemów agentowych umożliwia tworzenie inteligentnych systemów zdolnych do samodzielnego planowania, uczenia i interakcji z otoczeniem. W kontekście sektora finansowego oznacza to jakościowy skok – od reaktywnych chatbotów do autonomicznych agentów wspierających analitykę, zgodność regulacyjną i obsługę klienta. J. Kraprayoon, Z. Williams, R. Fayyaz, dz. cyt., s. 2–5; D. Leslie, C. Rincón, M. Briggs, A. Perini, S. Jayadeva, A. Borda, C. SJ. Burr Bennett, M. Aitken, S. Mahomed, J. Wong, M. Waller, C. Fischer, AI Explainability in Practice. The Alan Turing Institute 2024, s. 10–12.

65 J. Kraprayoon, Z. Williams, R. Fayyaz, dz. cyt., s. 1–3.

Reasumując, AI agenci to systemy zbudowane wokół dużych modeli językowych, wspierane przez tzw. scaffolding (m.in. pamięć długoterminową, planowanie, integrację z narzędziami cyfrowymi), które umożliwiają im autonomiczne wykonywanie złożonych zadań, komunikację z API, współpracę z innymi agentami i interakcję z użytkownikami końcowymi. Ich rozwój umożliwia realizację sekwencji działań bez potrzeby instrukcji krok po kroku[66].

W sektorze finansowym AI agenci mogą przynieść korzyści w kilku obszarach[67]:

1. Obsługa klienta (Customer Service 2.0):
    a. personalizacja kontaktu (agenci z pamięcią kontekstową i adaptacją do stylu klienta),
    b. wsparcie 24/7, z możliwością eskalacji do człowieka.
2. Compliance i raportowanie regulacyjne:
    a. automatyczne sprawdzanie dokumentów zgodnie z przepisami,
    b. przygotowanie draftów zgłoszeń do regulatorów.
3. Zarządzanie ryzykiem i analiza transakcji:
    a. wykrywanie anomalii z użyciem pamięci długoterminowej (co jest nieosiągalne dla klasycznych botów),
    b. ocena ryzyk operacyjnych czy reputacyjnych na podstawie dynamicznych danych (np. newsów, mediów społecznościowych).
4. Wsparcie dla doradców finansowych:
    a. przygotowanie spersonalizowanych analiz,
    b. interpretacja danych inwestycyjnych, predykcje rynkowe.

AI agenci reprezentują jedno z najbardziej obiecujących narzędzi transformacji technologicznej w sektorze finansowym. Ich zdolność do autonomicznego działania, planowania wieloetapowych procesów oraz integracji z systemami zewnętrznymi otwiera nowe możliwości w zakresie automatyzacji, analityki oraz zgodności regulacyjnej. Stanowią tym samym przyszłościowy komponent nowoczesnej architektury systemów bankowych i finansowych. Efektywne wdrożenie takich rozwiązań wymaga jednak spełnienia kilku kluczowych warunków.

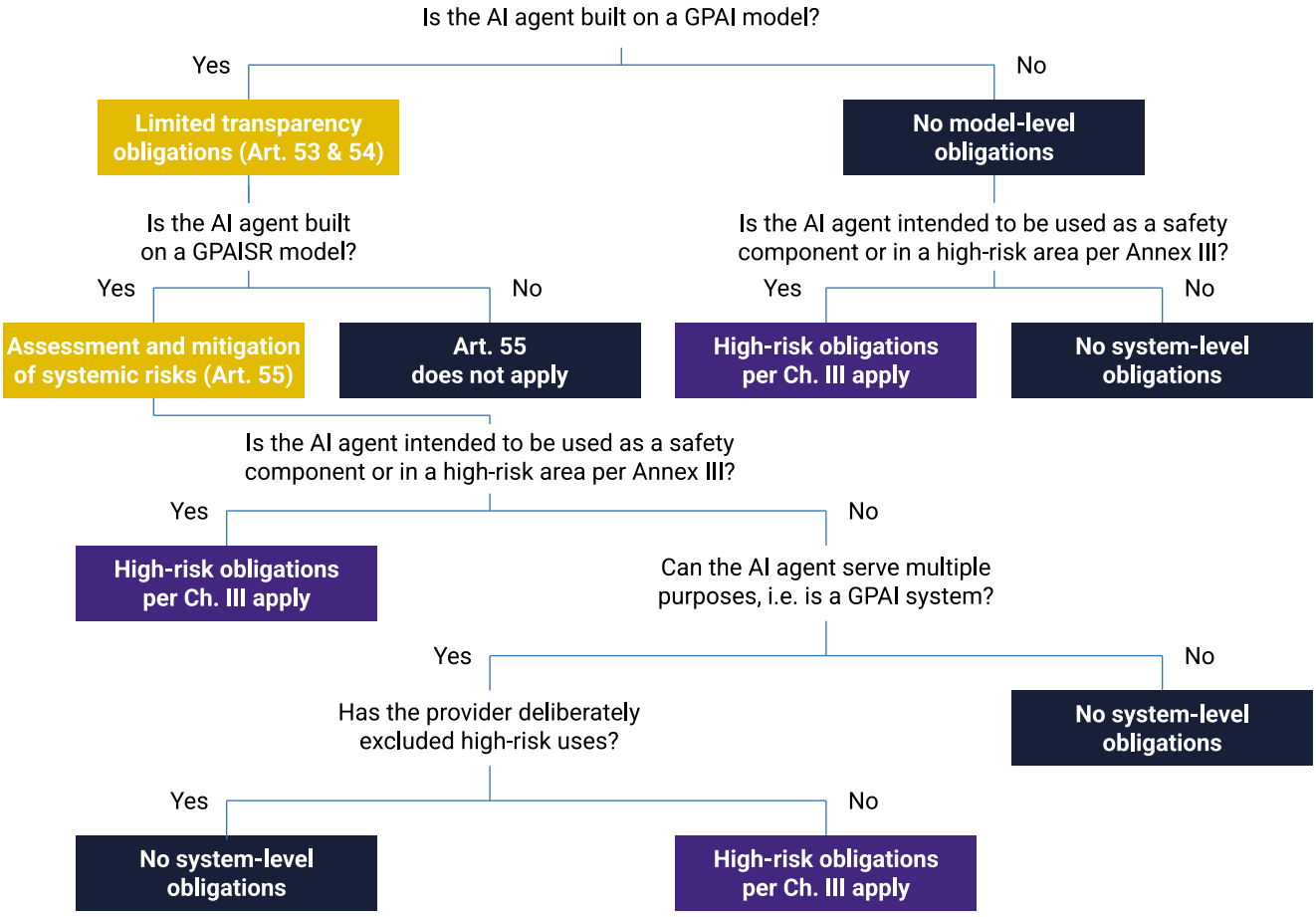
Po pierwsze, konieczna jest świadoma i przemyślana integracja agentów AI z istniejącymi systemami informatycznymi i bazami danych, przy uwzględnieniu logiki biznesowej i struktury procesów operacyjnych.

Po drugie, istotne jest ustanowienie przejrzystych zasad nadzoru i odpowiedzialności za działania systemów agentowych, szczególnie w kontekście wymogów wynikających z rozporządzenia AI Act (zob. rys.5) oraz norm ISO/IEC 42001 dotyczących zarządzania systemami AI.

Wreszcie, należy uwzględnić aspekty ekonomiczne, w tym rosnące koszty obliczeniowe związane z utrzymaniem agentów opartych na dużych modelach językowych, co wymaga optymalizacji i kontroli zużycia zasobów infrastrukturalnych. Właściwe zarządzanie tymi czynnikami może przesądzić o tym, czy AI agenci będą w organizacji źródłem innowacyjnej przewagi, czy ryzykownym i kosztownym eksperymentem[68] (zob. rys.6).

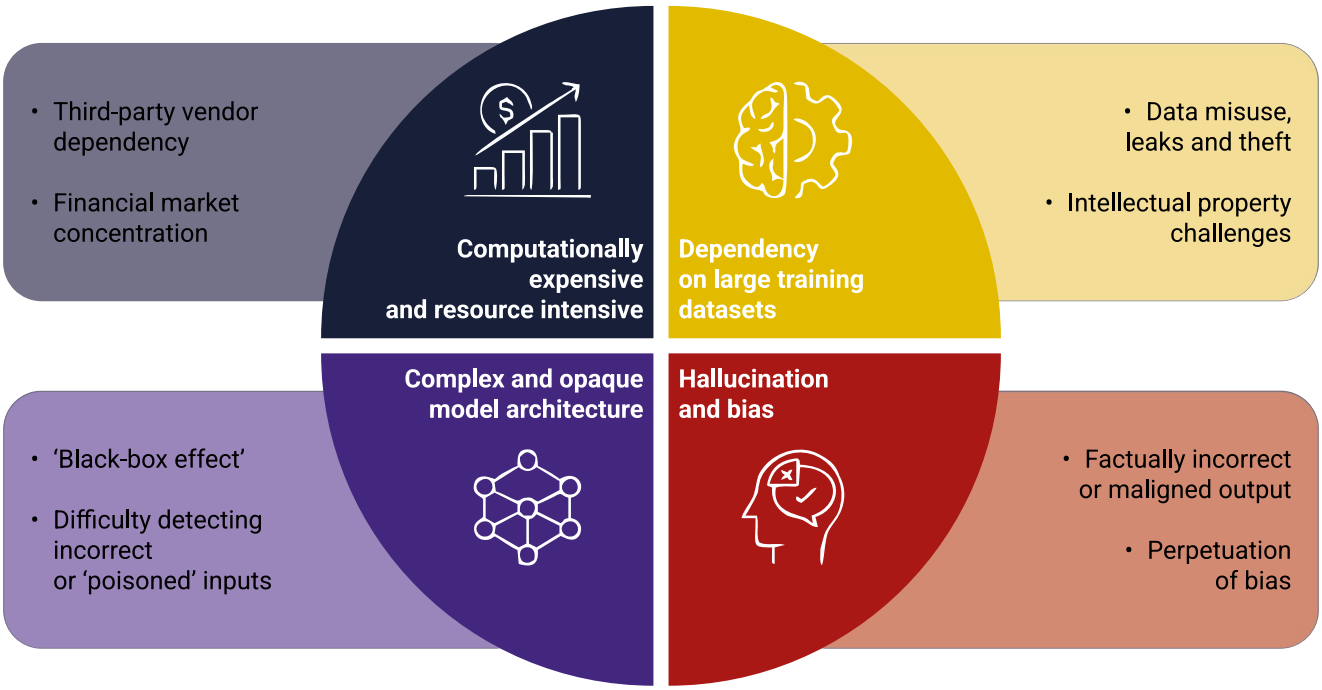
66 Tamże.
67 A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 85–90; J. Kraprayoon, Z. Williams, R. Fayyaz, dz. cyt., s. 2–4.
68 Ministry of Infrastructure and Water Management, dz. cyt., s. 5–28; S. Musch, M. Borrelli, C. Kerrigan, J. Bennison, dz. cyt., s. 5–10; J. Kraprayoon, Z. Williams, R. Fayyaz, dz. cyt., s. 3–6.

Rys. 5. Drzewo decyzyjne, które pozwala ustalić, kiedy Agent AI podlega pod AI Act.



Źródło: *A. Oueslati, R. States-Polet, Ahead od the Curve: Governing AI Agents Under the EU AI Act, The Future Society 2025, s. 27.*

Rys. 6. Koło ryzyk i wyzwań w wdrażaniu GenAI.



Źródło: *Hong Kong Institute for Monetary and Financial Research, Financial Services in the Era of Generative AI Facilitating Responsible Adoption, HKIMR Applied Research Report No.1/2025, s. 10.*

3. Personalizacja usług płatniczych

Personalizacja usług płatniczych polega na dostosowaniu interfejsów, funkcjonalności, komunikacji oraz rekomendacji produktów finansowych do indywidualnych potrzeb i zachowań użytkownika. Dzięki AI i uczeniu maszynowemu (ML) możliwe jest dynamiczne tworzenie profili użytkowników w oparciu o dane historyczne, kontekstowe i predykcyjne. Modele predykcyjne analizują m.in. częstotliwość transakcji, lokalizacje płatności, typy zakupów, preferencje urządzeń czy godziny aktywności klienta[69].

Na tej podstawie systemy AI mogą generować propozycje płatności ratalnych, przypomnienia o opłatach, sugerowane doładowania konta, a także dynamiczne limity płatności dopasowane do zdolności i historii użytkownika[70].

Do najczęściej wykorzystywanych narzędzi AI w personalizacji usług płatniczych należą[71]:

- Systemy rekomendacyjne, które na podstawie zachowań podobnych użytkowników (collaborative filtering) sugerują oferty kredytowe, modele płatności czy promocje kontekstowe;
- Segmentacja klientów oparta na klastrach, wykorzystywana do różnicowania komunikacji marketingowej, interfejsu aplikacji mobilnej oraz sposobu prezentowania produktów finansowych (np. automatyczne dobieranie układu menu czy „kafelków" z usługami);
- NLP (natural language processing) w ramach chatbotów i wirtualnych asystentów, które nie tylko rozumieją kontekst rozmowy, ale też dynamicznie personalizują treści – np. dostosowując ton wypowiedzi lub sugerując działania na podstawie wcześniejszych interakcji;
- Uczenie głębokie (deep learning), stosowane do przewidywania potrzeb płatniczych, ryzyk płynnościowych, a także automatycznego wnioskowania o potrzebie odroczenia płatności lub przesunięcia terminu.

Wykorzystanie tych narzędzi jest już standardem w największych instytucjach finansowych – np. Wells Fargo wykorzystuje AI do identyfikacji momentów życia klienta (np. narodziny dziecka, zmiana miejsca pracy) i proponowania adekwatnych rozwiązań płatniczych[72].

Przykładem udanej personalizacji usług płatniczych jest aplikacja banku Capital One, która oferuje klientom spersonalizowane analizy wydatków, automatyczne kategoryzowanie transakcji i predykcje przyszłych kosztów. Inny przykład to Revolut – fintech, który wdrożył uczenie maszynowe w module analityki finansowej, umożliwiając klientom otrzymywanie prognoz finansowych opartych na ich unikalnych wzorcach zachowań[73].

Banki i fintechy wykorzystują też generatywne modele językowe do komunikacji z klientami. Przykładem jest ChatGPT w aplikacji Klarna, który działa jako osobisty asystent płatniczy. Według firmy, jego wdrożenie pozwoliło na automatyzację obsługi klienta na poziomie odpowiadającym pracy ok. 700 etatów, bez obniżenia satysfakcji klienta[74].

69 Swiss Banking, dz. cyt, s. 22–24; A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 102–104.
70 A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 144–148.
71 Swiss Banking, dz. cyt, s. 22–23; A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 105–106, 108–111 oraz 173–174.
72 W publikacji „Generative AI in Banking Financial Services and Insurance" autorzy opisują przykład Wells Fargo i jego wykorzystania AI do tzw. „life event prediction", czyli identyfikacji ważnych momentów w życiu klienta i dostosowywania do nich oferty finansowej (m.in. systemów płatniczych i kredytowych). To klasyczny przykład zastosowania AI w hiperpersonalizacji usług finansowych. A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 106.
73 A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 110–111.
74 Raport firmy wskazuje, że ich AI asystent oparty na ChatGPT-4: obsłużył ponad 2/3 zapytań klientów w pierwszym miesiącu, osiągnął poziom obsługi odpowiadający 700 etatom, utrzymał poziom satysfakcji klientów na poziomie porównywalnym do wcześniejszej obsługi ludzkiej. Klrana, Klarna AI assistant handles two-thirds of customer service chats in its first month, 2024, https://www.klarna.com/international/press/klarna-ai-assistant-handles-two-thirds-of-customer-service-chats-in-its-first-month/

Personalizacja przynosi wiele korzyści dla instytucji finansowych[75]:

- zwiększenie retencji klientów i redukcja ich migracji do konkurencji;
- możliwość prowadzenia hiperprecyzyjnego marketingu i zwiększenia konwersji;
- redukcja kosztów obsługi klienta dzięki automatyzacji;
- wczesne wykrywanie problemów płynnościowych i oferowanie proaktywnych rozwiązań;
- lepsza zgodność z regulacjami dotyczącymi przejrzystości i uczciwego traktowania klienta.

Ponadto, personalizacja wspierana AI wzmacnia wartość marki i postrzeganą innowacyjność instytucji, co ma znaczenie zarówno wśród klientów indywidualnych, jak i inwestorów[76].

Zastosowanie AI w personalizacji wymaga ostrożności w zakresie ochrony prywatności, przejrzystości działania algorytmów oraz unikania uprzedzeń (bias). Szczególnie wrażliwe są tu zagadnienia zgodności z RODO i obowiązkiem wyjaśnienia klientowi, na jakiej podstawie podejmowane są decyzje np. o limitach czy propozycjach finansowych. Należy również unikać tzw. „filter bubbles", czyli sytuacji, w której klient otrzymuje tylko ograniczony zestaw opcji na podstawie zbyt wąskiego profilu behawioralnego[77].

Z uwagi na rosnące wymagania klientów i rozwój technologii, personalizacja usług płatniczych będzie jednym z głównych trendów w bankowości cyfrowej w nadchodzących latach. Rozwój modeli multimodalnych (łączących analizę tekstu, głosu i obrazu), integracja z urządzeniami IoT oraz nowe interfejsy (np. głosowe) jeszcze bardziej zwiększą precyzję i intuicyjność personalizowanych usług[78].

Banki, które zainwestują w silniki personalizacji oparte na AI i jednocześnie zapewnią zgodność z regulacjami i etyką, będą miały znaczną przewagę konkurencyjną na zglobalizowanym rynku usług płatniczych[79].

3.1. Algorytmy AI w analizie zachowań użytkowników

Współczesne systemy bankowości cyfrowej opierają się coraz częściej na analizie behawioralnej klientów, prowadzonej w sposób ciągły i nienarzucający się. Dzięki algorytmom AI instytucje finansowe mogą nie tylko klasyfikować użytkowników na podstawie danych deklaratywnych czy transakcyjnych, ale również rozpoznawać subtelne wzorce w sposobie korzystania z usług. Obejmuje to analizę czasu reakcji w aplikacji, nawyków zakupowych, zachowań w interfejsie oraz mikrointerakcji z systemem – dane te tworzą tzw. ślad behawioralny użytkownika[80].

Kluczową rolę w tej analizie odgrywają **modele sekwencyjne** (np. LSTM, GRU), zdolne do wychwytywania długoterminowych zależności między działaniami użytkownika a jego intencjami płatniczymi. Przykładowo, wcześniejsze wzorce zakupowe powiązane z sezonowością, nagłymi wzrostami kosztów lub zmianami lokalizacji, mogą sygnalizować potrzebę dostosowania oferty kredytowej lub zabezpieczenia limitów płatniczych[81].

Równolegle stosowane są modele **klasteryzacji adaptacyjnej**, które w odróżnieniu od statycznych profili demograficznych, grupują klientów w czasie rzeczywistym na podstawie ich aktywności. Zastosowanie algorytmów takich jak k-means++ czy DBSCAN pozwala na identyfikację mikrosegmentów użytkowników, których potrzeby i styl korzystania z produktów finansowych różnią się w zależności od fazy życia,

75 A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 112–113.
76 Tamże, s. 113.
77 W raporcie „Generative AI in Banking – A Comprehensive Overview" opisano zagrożenie tworzenia zbyt hermetycznych rekomendacji w systemach AI, prowadzące do zjawiska „AI-induced personalization traps", które mogą ograniczać świadomy wybór klienta i obniżać jego poczucie autonomii decyzyjnej. To właśnie odpowiada koncepcji „filter bubble" w kontekście bankowości i płatności. Swiss Banking, dz. cyt, s. 30.
78 Swiss Banking, dz. cyt, s. 33–34.
79 Autorzy książki „Generative AI in Banking Financial Services and Insurance" jednoznacznie wskazują, że instytucje finansowe, które połączą personalizację z odpowiedzialnym zarządzaniem danymi, przejrzystością i zgodnością regulacyjną, osiągną przewagę konkurencyjną i będą lepiej przygotowane na rosnące oczekiwania klientów oraz nadzorców. A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 113–114.
80 W szczególności chodzi o znaczenie danych behawioralnych i dynamicznej klasyfikacji klientów nie tylko na podstawie danych statycznych, ale także na podstawie obserwacji wzorców korzystania z systemów bankowych. A. Gupta, D. N. Dwivedi, J. Shah, Artificial Intelligence Applications in Banking and Financial Services, Springer 2023, s. 79 i n.
81 Tamże, s. 131.132.

stanu konta czy zachowań kontekstowych (np. lokalizacja, pora dnia, typ urządzenia)[82].

Jednym z bardziej zaawansowanych podejść jest tzw. **behavioral embedding** – technika reprezentowania użytkownika jako wektora w przestrzeni wielowymiarowej, wygenerowanego przez głęboką sieć neuronową na podstawie jego historii interakcji. Dzięki temu systemy mogą precyzyjnie przewidywać nie tylko najbliższe działania, ale i długoterminowe preferencje zakupowe czy finansowe. Takie podejście znajduje zastosowanie np. w generowaniu ofert uprzedzających: proponowaniu odroczenia płatności zanim klient sam rozważy taką opcję, czy przedstawianiu oszczędnościowego planu ratalnego w reakcji na pojawiające się schematy nadmiernych wydatków[83].

Innowacją staje się również **kontekstualne rozpoznawanie anomalii** – polegające nie na odrzucaniu niestandardowych działań klienta, lecz na ich analizie w szerszym kontekście czasowym i transakcyjnym. Przykładowo, pojedynczy zakup o wysokiej wartości w nietypowym miejscu nie zostanie automatycznie oznaczony jako podejrzany, jeśli system uwzględni wcześniejszą rezerwację hotelu lub zmianę geolokalizacji użytkownika. W takich przypadkach AI działa jako filtr ryzyka dostosowany do indywidualnego profilu klienta[84].

Zastosowanie generatywnych modeli językowych (LLM) w analizie behawioralnej umożliwia też bardziej spersonalizowaną interpretację zachowań – np. analizę treści zapytań klienta kierowanych do asystenta AI w połączeniu z jego aktywnością transakcyjną. Integracja NLP i uczenia nienadzorowanego pozwala tworzyć narracyjne profile użytkowników, w których identyfikuje się nie tylko ich potrzeby, ale i emocje oraz motywacje zakupowe[85].

Ostatecznie skuteczność algorytmów analizy behawioralnej zależy nie tylko od ich matematycznej precyzji, ale także od otoczenia danych, w jakim funkcjonują. Jak podkreśla Swiss Bankers Association, zarządzanie ryzykiem modeli AI musi być nierozerwalnie związane z jakością danych wejściowych, ich integralnością oraz zgodnością ze standardami etycznymi i regulacyjnymi. Dotyczy to zwłaszcza sytuacji, w których dane behawioralne mogą prowadzić do niejawnych klasyfikacji użytkowników i wpływać na dostęp do usług płatniczych – stąd rosnące znaczenie audytów algorytmicznych oraz mechanizmów zapewniających explainability i auditability modeli[86].

3.2. Dynamiczne modele oceny zdolności kredytowej

Klasyczne modele oceny zdolności kredytowej, oparte głównie na danych deklaratywnych, wskaźnikach finansowych oraz punktacji historycznej, stają się niewystarczające w złożonym środowisku danych i oczekiwań klientów XXI wieku. Sztuczna inteligencja (AI), a w szczególności uczenie maszynowe (ML), umożliwia przejście od statycznej oceny do **dynamicznego, ciągłego monitorowania zdolności kredytowej w czasie rzeczywistym**, uwzględniającego szeroki kontekst behawioralny, transakcyjny i predykcyjny[87].

W tradycyjnych systemach scoringowych, np. FICO czy BIK, decyzja kredytowa opiera się na ograniczonym zestawie wskaźników i punktów. W odróżnieniu od nich, nowoczesne modele AI analizują setki, a nawet tysiące zmiennych – od danych o płatnościach stałych, przez fluktuacje wydatków, po zachowania mobilne i informacje kontekstowe, takie jak lokalizacja czy pora dnia transakcji[88].

Modele te są zdolne do adaptacji, co oznacza, że w czasie trwania relacji kredytowej mogą reagować na zmieniające się warunki klienta – np. przewidywać trudności finansowe na podstawie wzrostu zadłużenia lub spadku dochodów jeszcze zanim dojdzie do opóźnień w spłatach[89].

82 Tamże, 128–129.
83 A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 104–105.
84 Tamże, s. 111.
85 Tamże, 108–109.
86 Swiss Banking, dz. cyt, s. 23–24.
87 Na przykład autorzy A. Gupta, D. N. Dwivedi, J. Shah wskazują na ograniczenia klasycznych metod scoringowych i przedstawiają AI jako fundament nowoczesnych, dynamicznych systemów oceny kredytowej, integrujących dane z wielu źródeł i reagujących w czasie rzeczywistym na zmiany w zachowaniu finansowym klientów. A. Gupta, D. N. Dwivedi, J. Shah, dz. cyt, s. 123–124.
88 A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 115.
89 Warto podkreślić, że autorzy A. Gupta, D. N. Dwivedi, J. Shah wyjaśniają, że jedną z kluczowych przewag AI w ocenie zdolności kredytowej jest jej zdolność adaptacyjna, która pozwala reagować na dynamiczne zmiany sytuacji klienta,

Coraz częściej wykorzystywane są **alternatywne dane** (alternative data), obejmujące m.in. historię opłat za media, aktywność w mediach społecznościowych, dane z urządzeń IoT, a także analizę semantyczną kontaktu z bankiem (np. treść zapytań do chatbota). Tego typu dane zwiększają inkluzywność systemu kredytowego, pozwalając ocenić zdolność klienta, który nie posiada historii kredytowej w tradycyjnym rozumieniu[90].

Algorytmy ML, takie jak **random forest**, **gradient boosting** czy **deep neural networks**, umożliwiają tworzenie nieliniowych, warstwowych modeli scoringowych, które z wysoką dokładnością klasyfikują ryzyko kredytowe i umożliwiają indywidualne podejście do klienta[91].

Wprowadzenie dynamicznych modeli scoringowych musi iść w parze z wymogami przejrzystości, które stają się obligatoryjne w świetle rozporządzenia **AI Act** oraz przepisów takich jak RODO. Systemy AI stosowane do oceny zdolności kredytowej są klasyfikowane jako **high-risk**, co nakłada obowiązek zapewnienia ich wyjaśnialności (explainability), rzetelności i odporności na uprzedzenia (bias)[92].

Rozwiązaniem tego problemu są techniki **explainable AI (XAI)** – np. LIME czy SHAP – które pozwalają prześledzić, jakie zmienne miały wpływ na decyzję kredytową i udostępnić klientowi logiczne uzasadnienie tej decyzji[93].

Nowoczesne systemy scoringowe implementują koncepcję **continuous credit assessment** – oceny zdolności kredytowej nie tylko w momencie składania wniosku, ale także przez cały okres trwania relacji klienta z instytucją finansową. Dzięki temu możliwe jest m.in.[94]:

- automatyczne dopasowanie limitu kredytowego do bieżącej sytuacji użytkownika,
- wczesne ostrzeganie o ryzyku niewypłacalności,
- dynamiczne zmiany warunków umowy kredytowej w zależności od zmian w zachowaniu klienta.

Banki takie jak **JPMorgan Chase**, **Capital One** i fintechy typu **Tink** czy **Upstart** wdrażają już takie systemy, integrując dynamiczne scoringi z całym ekosystemem zarządzania ryzykiem i zgodnością[95].

Dynamiczne modele scoringowe wymagają nie tylko nadzoru technicznego, ale także nowej kultury zarządzania ryzykiem – łączącej kompetencje data science z wiedzą prawną, etyczną i nadzorczą. W przeciwnym razie, nawet bardzo dokładne algorytmy mogą generować decyzje trudne do uzasadnienia lub wręcz dyskryminujące. Dlatego tak istotne jest przeprowadzanie **testów odporności modeli**, regularnych audytów i weryfikacji działania systemu przez niezależnych ekspertów[96].

3.3. Rekomendacje i oferty dostosowane do profilu klienta

Jednym z kluczowych zastosowań sztucznej inteligencji w nowoczesnych systemach płatniczych i bankowości cyfrowej jest generowanie rekomendacji produktowych i ofert finansowych dopasowanych do indywidualnego profilu użytkownika. AI umożliwia wyjście poza tradycyjne podejście segmentacyjne i pozwala na formułowanie propozycji usług opartych na bieżących potrzebach, zachowaniach oraz kontekście działania klienta.

Silniki rekomendacyjne wykorzystywane przez instytucje finansowe opierają się na technikach takich jak collaborative filtering, content-based filtering oraz coraz częściej – uczenie głębokie i embeddingi behawioralne. Modele te analizują m.in. częstotliwość

co umożliwia proaktywne zarządzanie ryzykiem kredytowym i zwiększa skuteczność wczesnych interwencji. A. Gupta, D. N. Dwivedi, J. Shah, dz. cyt, s. 126–127.

90 Alternative data pozwalają ocenić zdolność kredytową klientów wcześniej wykluczonych z systemu finansowego, zwiększając inkluzywność kredytową. Wśród przykładów wymieniają m.in. dane z rachunków domowych, aktywność online, dane IoT oraz analizę konwersacyjną (NLP) zapytań klienta. A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 116–117.

91 A. Gupta, D. N. Dwivedi, J. Shah, dz. cyt, s. 125–126.

92 Na przykład raport Swiss Bank wyjaśnia, że zgodnie z AI Act systemy AI używane do oceny zdolności kredytowej są objęte kategorią high-risk i muszą spełniać kryteria przejrzystości, odporności na błędy i uprzedzenia, a także umożliwiać audytowalność i wyjaśnialność działania algorytmu w odniesieniu do decyzji finansowych.Swiss Banking, dz. cyt, s. 26–27.

93 A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 118.

94 A. Gupta, D. N. Dwivedi, J. Shah, dz. cyt, s. 127–128.

95 A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 117–118.

96 Swiss Banking, dz. cyt, s. 26–27; S. Musch, M. Borrelli, C. Kerrigan, J. Bennison, dz. cyt., s. 5–6.

i rodzaj transakcji, wzorce zakupowe, aktywność na urządzeniach mobilnych, lokalizacje geograficzne oraz dane z wcześniejszej interakcji z systemem bankowym lub asystentem AI[97].

Zastosowanie takich technik pozwala instytucjom finansowym generować rekomendacje w czasie rzeczywistym, np.[98]:

- propozycje dopasowanych limitów kredytowych lub kart z funkcją cashback w kategoriach najczęściej używanych przez klienta;
- przypomnienia o nadchodzących płatnościach lub automatyczne sugestie ich rozłożenia na raty;
- dynamiczne oferty oszczędnościowe lub inwestycyjne oparte na analizie bieżącej zdolności płatniczej i stylu wydatków użytkownika.

Od rekomendacji transakcyjnych do doradztwa proaktywnego AI umożliwia przejście od prostych sugestii do proaktywnego doradztwa finansowego, w którym systemy nie tylko reagują na potrzeby klienta, ale także antycypują je. Przykładem może być wykrycie rosnących kosztów w konkretnej kategorii (np. transport) i zaproponowanie alternatywnej usługi lub planu wydatków. Coraz częściej instytucje wdrażają modele generatywne, które potrafią syntetyzować komunikaty w formie zrozumiałej dla klienta i prowadzić z nim dialog oparty na historii jego działań[99].

Systemy takie wykorzystywane są m.in. przez Capital One (Eno), Bank of America (Erica) oraz Klarna, której chatbot generatywny odpowiada za większość interakcji z klientem i może pełnić funkcję osobistego doradcy płatniczego[100].

Dobrze zaprojektowane systemy rekomendacyjne nie tylko zwiększają satysfakcję klienta, ale także wpływają na jego decyzje finansowe, poprawiając np. terminowość spłat lub skłonność do oszczędzania. AI umożliwia także wdrożenie tzw. micro-personalization, czyli oferty finansowej skalowanej do poziomu jednostkowego użytkownika – nie tylko w treści, ale i formie prezentacji (np. kolorystyka aplikacji, ton komunikacji, typ interfejsu)[101].

Generowanie ofert finansowych przy użyciu AI wymaga przestrzegania zasady przejrzystości oraz unikania nadużyć behawioralnych. Rekomendacje nie mogą wprowadzać w błąd ani opierać się na manipulacji percepcyjnej (tzw. dark patterns). Konieczne jest również zapewnienie zgodności z RODO – w szczególności art. 22, który dotyczy decyzji opartych wyłącznie na zautomatyzowanym przetwarzaniu[102].

Zgodnie z AI Act oraz standardem ISO/IEC 42001, oferowanie usług w sposób personalizowany musi uwzględniać prawo klienta do uzyskania uzasadnienia decyzji oraz możliwość jej zakwestionowania oraz informacji że decyzja została podjęta przez system AI. Oznacza to potrzebę wdrażania mechanizmów explainable AI (XAI) w systemach rekomendacyjnych, w szczególności w przypadku ofert związanych z kredytami, limitami zadłużenia lub inwestycjami wysokiego ryzyka[103].

97 A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 103–104.
98 Tamże, 105–106.
99 Tamże, 107–108.
100 Swiss Banking, dz. cyt, s. 34–35.
101 A. Saxena, S. Verma, J. Mahajan, dz. cyt., s. 109–110.
102 Swiss Banking, dz. cyt, s. 29–30.
103 Swiss Banking, dz. cyt, s. 26–27; S. Musch, M. Borrelli, C. Kerrigan, J. Bennison, dz. cyt., s. 5–6.

4. Ryzyko algorytmiczne i zagrożenia związane z AI w płatnościach

Wykorzystanie sztucznej inteligencji (AI) w systemach płatniczych i szerszym sektorze usług finansowych przynosi potencjalny szereg innowacyjnych usprawnień[104], ale również rodzi nowe formy ryzyka i zagrożeń. Obejmują one ryzyko algorytmiczne, cyberzagrożenia, ryzyko systemowe, ryzyko dyskryminacji oraz problemy związane z przejrzystością i odpowiedzialnością.

4.1. Dyskryminacja algorytmiczna i brak przejrzystości decyzji AI

Systemy uczenia maszynowego wykorzystywane w finansach są narażone na degradację wydajności modeli w czasie. Modele te, trenowane na danych historycznych, mogą ulegać dezaktualizacji wskutek zmian w zachowaniach użytkowników, strukturze rynków lub w wyniku intencjonalnych manipulacji danymi wejściowymi. Przykładowo, zmiana wzorców zachowań płatniczych może prowadzić do fałszywych odrzuceń transakcji lub błędnego przypisania poziomu ryzyka kredytowego[105].

Algorytmy mogą prowadzić do niezamierzonej dyskryminacji użytkowników na etapie klasyfikacji ryzyka, oceny zdolności kredytowej czy zatwierdzania transakcji. Dyskryminacja ta może mieć charakter bezpośredni (np. w oparciu o zakodowaną cechę demograficzną) lub pośredni, gdy model bazuje na zmiennych skorelowanych z cechami chronionymi, np. kodem pocztowym lub historią uczelni[106].

Badania wykazują, że osoby z uczelni reprezentujących mniejszości etniczne otrzymywały gorsze oferty kredytowe, mimo porównywalnych parametrów ekonomicznych[107]. W sektorze płatności może to skutkować ograniczonym dostępem do usług finansowych lub wyższymi kosztami kredytowania.

Wiele modeli AI, zwłaszcza opartych na głębokim uczeniu (deep learning), działa w sposób nieprzejrzysty nawet dla ich twórców. Problem „czarnej skrzynki" uniemożliwia dokładne zrozumienie, dlaczego system podjął określoną decyzję – np. odrzucił transakcję lub przyznał niski rating kredytowy. Brak możliwości uzyskania wyjaśnienia decyzji narusza prawo do skutecznego środka odwoławczego oraz zagraża zaufaniu społecznemu[108].

Zgodnie z wymogami m.in. RODO, osoby fizyczne mają prawo do uzyskania „znaczących informacji o zasadach podejmowania decyzji" w przypadku zautomatyzowanego przetwarzania[109]. W praktyce jednak stosowanie tych zasad w kontekście systemów opartych na AI wciąż jest ograniczone.

Rozporządzenie AI Act oraz liczne wytyczne (np. ENISA, IEEE, ISO) wskazują konieczność wdrożenia mechanizmów wykrywania i redukcji uprzedzeń, zapewnienia przejrzystości decyzji oraz wprowadzenia obowiązkowych ocen wpływu na prawa podstawowe (Fundamental Rights Impact Assessments, FRIA)[110]. ENISA w swoim wielowarstwowym modelu bezpieczeństwa AI wskazuje potrzebę komplementarnego podejścia: od ogólnego bezpieczeństwa

104 Autorzy pisali o tym w innych raportach, wraz z M. Nowakowski i G. Bar. Zob. D. Szostek, M. Nowakowski, G. Bar, R. Prabucki, *Zastosowanie sztucznej inteligencji w bankowości – szanse oraz zagrożenia*, Fundacja Warszawski Instytut Bankowości, Warszawa 2022.
105 W. Spiess-Knafl, Artificial Intelligence and Blockchain for Social Impact, Routledge 2023, s.
106 C. Kerrigan, Artificial Intelligence. Law and Regulation, Elgar Publishing 2022, s. 422–427.
107 L. Naudts, A. Vedder, Fairness and Artificial Intelligence [w:] N. A. Smuha (red.), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*, Cambridge University Press 2025, s. 79 i n.
108 Tamże, s. 83–100.; L. Lauwaert, A. Oimann, *Moral Responsibility and Autonomous Technologies: Does AI Face a Responsibility Gap?* [w:] N. A. Smuha (red.), dz. cyt, s. 101 i n.
109 Por. art. 22 RODO oraz omówienie w C. Kerrigan, dz. Cyt., s. 163–172.
110 P. Hense, T. Mustać, *AI Act Compact. Compliance, managment & use cases in corporate practice*, Deutscher Fachverlag GmbH, Fachmedien Recht und Wirtschaft, Frankfurt am Main 2025, s. 179–204.

infrastruktury ICT, przez specyfikę systemów AI, po praktyki sektorowe – np. w finansach[111].

AI w systemach płatniczych staje się nieodłącznym elementem transformacji cyfrowej, ale jednocześnie wymaga zaawansowanych mechanizmów zarządzania ryzykiem. Kluczowe pozostają: zapewnienie zgodności z regulacjami dotyczącymi przejrzystości i niedyskryminacji, wdrożenie audytowalnych i wyjaśnialnych modeli decyzyjnych oraz ciągły nadzór nad działaniem systemów AI w środowisku rzeczywistym.

4.2. Ryzyko cyberzagrożeń i oszustw z wykorzystaniem AI

Sektor bankowy jest jednym z najbardziej narażonych na zagrożenia w cyberprzestrzeni obszarów, co wynika z jego znaczenia infrastrukturalnego, intensywnego wykorzystania technologii cyfrowych oraz potencjalnych korzyści finansowych dla sprawców ataków. W niniejszym rozdziale analizujemy aktualne cyberzagrożenia dla instytucji bankowych w Europie, ze szczególnym uwzględnieniem danych z raportu ENISA dla sektora finansowego (2023–2024) oraz wyzwań wynikających z wykorzystania sztucznej inteligencji (AI) – zarówno jako narzędzia ataku, jak i celu. Przedstawiono główne typy incydentów, ich przyczyny i konsekwencje oraz rekomendacje dotyczące poprawy odporności sektora.

Sztuczna inteligencja pełni podwójną rolę w kontekście cyberzagrożeń: jako narzędzie przestępcze (np. generowanie wiadomości phishingowych[112], deepfake[113]) oraz jako cel ataku (np. ataki adversarialne, data poisoning)[114].

4.2.1. Incydenty bezpieczeństwa, a AI

Bankowość (finanse) jako sektor o krytycznym znaczeniu dla gospodarki, podlega nieustannej presji adaptacyjnej w kontekście transformacji cyfrowej. Rozwój usług zdalnych, otwartej bankowości oraz wykorzystanie AI prowadzą do zwiększenia powierzchni ataku i podatności na nowe wektory zagrożeń. Jak pokazuje raport ENISA, sektor ten był trzecim najczęściej atakowanym w UE w okresie od stycznia 2023 r. do czerwca 2024 r., ustępując jedynie administracji publicznej i transportowi[115].

W analizowanym okresie odnotowano 301 incydentów cybernetycznych wymierzonych bezpośrednio w banki – co stanowi 46% wszystkich zdarzeń w sektorze finansowym. Najczęstsze rodzaje incydentów to: ataki DDoS[116], naruszenia danych, ransomware oraz zaawansowane malware, takie jak trojany bankowe[117].

Zidentyfikowane podatności, takie jak niewystarczające zabezpieczenia w łańcuchu dostaw IT, błędy w konfiguracji systemów czy brak świadomości pracowników i klientów, mają istotne implikacje dla bezpieczeństwa systemów sztucznej inteligencji stosowanych w bankowości.

W pierwszej kolejności należy zauważyć, że współczesne rozwiązania AI w instytucjach finansowych

111 ENISA, Multilayer Framework for Good Cybersecurity Practices for AI, 2023, s. 6–34.
112 Phishing, zgodnie za ENISA, jest to: „oszukańcza próba kradzieży danych użytkownika, takich jak dane logowania, informacje o karcie kredytowej, a nawet pieniądze, przy użyciu technik inżynierii społecznej. Ten rodzaj ataku jest zwykle przeprowadzany za pośrednictwem wiadomości e-mail, które wydają się być wysyłane z renomowanego źródła, z zamiarem przekonania użytkownika do otwarcia złośliwego załącznika lub podążania za fałszywym adresem URL. Ukierunkowana forma phishingu zwana „spear phishing" polega na wcześniejszym badaniu ofiar, aby oszustwo wydawało się bardziej autentyczne, co czyni go jednym z najbardziej skutecznych rodzajów ataków na sieci przedsiębiorstw". ENISA, Phishing, Chalandri 2020, s. 2 i n.
113 Deepfake, czy jak też przetłumaczono to w wyroku Sądu Najwyższego – przerysowana prezentacja, to zjawisko, w którym autor nie prezentuje rzeczywistości. Jest to więc treść nieprawdziwa, w której o prawnym zakresie decyduje cel autora, który taką treść przygotował. Jest to ważne, bo choć przerysowana prezentacja nie ukazuje rzeczywistości, to może zawierać prawdziwe twarze, miejsca, znaki handlowe itp. Zob. Wyrok Sądu Najwyższego – Izba Cywilna z dnia 11 sierpnia 2022 r. II CSKP 313/22.
114 K. Zieliński, A. Ślusarek, Wykorzystanie sztucznej inteligencji jako zagrożenie dla klientów rynku finansowego [w:] A. Szczęsna, M. Stachoń (red.), *Cyberbezpieczeństwo AI. AI w Cyberbezpieczeństwie*, wyd. NASK – Państwowy Instytut Badawczy, Warszawa 2023, s. 82 i n.
115 ENISA, Threat landscape: finance sector, Chalandri 2025, s. 4 i n.
116 Zgodnie za ENISA, DoS to „ (…) ataki na dostępność, w których atakujący częściowo lub całkowicie utrudniają legalne korzystanie z usługi celu poprzez wyczerpywanie lub wykorzystywanie zasobów celu przez pewien okres czasu.". ENISA, *Threat landscape for DoS attacks*, Chalandri 2023, s. 6.
117 Tamże.

często bazują na zewnętrznych komponentach – bibliotekach programistycznych, środowiskach chmurowych czy dostawcach modeli uczenia maszynowego. Taka zależność sprawia, że nieadekwatne zabezpieczenia w łańcuchu dostaw mogą prowadzić do poważnych zagrożeń, takich jak wstrzyknięcie złośliwego kodu do modelu, manipulacja danymi treningowymi (tzw. data poisoning) czy tworzenie ukrytych „tylnych furtek" umożliwiających przejęcie kontroli nad systemem. ENISA podkreśla, że tego rodzaju zagrożenia są szczególnie aktualne w przypadku wdrożeń AI w środowiskach złożonych i silnie zależnych od podmiotów trzecich[118].

Kolejnym istotnym aspektem są błędy konfiguracyjne, które w środowiskach wykorzystujących z AI mogą skutkować nieautoryzowanym dostępem do interfejsów programistycznych (API), ujawnieniem danych wejściowych lub wyjściowych modelu, a nawet eksfiltracją parametrów samego modelu. Takie uchybienia mogą umożliwić przeprowadzenie ataków polegających na odtworzeniu danych wejściowych (tzw. model inversion) lub na wprowadzeniu specjalnie spreparowanych danych wejściowych, które zakłócą działanie algorytmu[119].

Nie można również pominąć czynnika ludzkiego. Niewystarczająca świadomość pracowników, którzy obsługują lub nadzorują systemy AI, może prowadzić do niezauważenia symptomów błędnego „uczenia się" modeli, występowania uprzedzeń decyzyjnych (bias) czy powielania nieprawidłowych schematów działania. Z kolei brak świadomości klientów, którym AI może być przedstawiane jako element interfejsu (np. w postaci inteligentnych asystentów czy chatbotów) sprawia, że są oni bardziej podatni na manipulacje, takie jak oszustwa phishingowe czy ataki typu deepfake, bazujące na syntetycznych obrazach i głosach[120].

Wszystkie te czynniki – infrastrukturalne, technologiczne i ludzkie – stanowią krytyczne elementy ryzyka dla systemów sztucznej inteligencji i muszą być uwzględniane w analizie zagrożeń oraz w praktyce zarządzania bezpieczeństwem w sektorze bankowym.

Wśród najczęstszych skutków incydentów w bankowości wymienia się straty finansowe, uszkodzenie reputacji, zakłócenia operacyjne oraz ujawnienie danych o wysokiej wartości. Skutki te nabierają dodatkowego znaczenia w kontekście wykorzystywania sztucznej inteligencji, której obecność w procesach bankowych nie tylko zwiększa efektywność, ale jednocześnie generuje nowe, specyficzne ryzyka[121].

Utrata reputacji, tradycyjnie związana z wyciekiem danych lub przerwą w świadczeniu usług, może być spotęgowana przez incydenty, w których zawiodły mechanizmy oparte na AI. Opinie publiczne są szczególnie wrażliwe na doniesienia o „samodzielnych decyzjach" algorytmów, które skutkują odmową kredytu, niesłuszną blokadą konta czy ujawnieniem prywatnych danych. W kontekście AI, reputacyjne ryzyko może również wynikać z przypadków dyskryminacji (np. ze względu na płeć lub pochodzenie etniczne), które wynikają z błędnie wytrenowanych modeli[122].

Zakłócenia operacyjne wywołane incydentami mogą objąć nie tylko systemy transakcyjne, ale również wewnętrzne moduły decyzyjne oparte na AI. Przerwanie działania takich systemów wpływa na zdolność banku do obsługi klientów, podejmowania decyzji kredytowych, zarządzania ryzykiem czy wykrywania nadużyć. Co więcej, modele AI wymagają ciągłego dostępu do danych i infrastruktury obliczeniowej – ich niedostępność może całkowicie unieruchomić niektóre procesy biznesowe[123].

118 ENIS, *Multilayer Framework for Good Cybersecurity Practices for Artificial Intelligence.* European Union Agency for Cybersecurity, Chalandri 2023, s. 17–19, zob. też. S. Moseley, *Automating Deception AI's Evolving Role in Romance Fraud*, The Alan Turing Institute 2025, s. 3 i n.

119 Opisane ryzyka są ujęte w warstwie „AI-specific cybersecurity practices" modelu wielowarstwowego ENISA i odnoszą się do podatności charakterystycznych dla wdrożeń modeli uczenia maszynowego w systemach realnego świata. ENISA, *Multilayer Framework for Good Cybersecurity Practices for Artificial Intelligence.* European Union Agency for Cybersecurity, dz. cyt., s. 20–22, zob. też. S. Moseley, dz. cyt., s. 3 i n.

120 K. Zieliński, A. Ślusarek, dz. cyt., s. 81–83.

121 Fragment ten choć pochodzi z raportu ENISA, to dobrze wpisuje się także we wnioski z raportu AI and Serious Online Crime (CETaS), który wskazuje na systemowe konsekwencje AI-enabled crime – w tym finansowe i reputacyjne szkody dla instytucji. ENISA, *Threat landscape: finance sector*, dz. cyt., s. 10–13; J. Burton, A. Janjeva, S. Moseley, A. Moseley, *AI and Serious Crime*, The Alan Turing Institute 2025, s. 4 i n.

122 K. Silicki, Cyberbezpieczeństwo systemów wykorzystujących sztuczną inteligencję w świetle raportów ENISA [w:] A. Szczęsna, M. Stachoń (red.), *Cyberbezpieczeństwo AI. AI w Cyberbezpieczeństwie*, wyd. NASK – Państwowy Instytut Badawczy, Warszawa 2023, s. 10 i n.

123 ENISA, *Threat Landscape: Finance Sector*, dz. cyt., s. 13–15; ENISA, *Multilayer Framework for Good Cybersecurity Practices for AI*, dz. cyt., s. 25–26.

Ujawnienie danych o wysokiej wartości, takich jak dane treningowe, zbiory walidacyjne, wewnętrzne parametry modeli lub wrażliwe dane klientów używane jako wejścia do systemów AI, stanowi szczególne zagrożenie. Wyciek takich informacji umożliwia przeprowadzenie tzw. ataków reidentyfikacyjnych, rekonstrukcję zachowań decyzyjnych modelu lub jego kopiowanie przez konkurencję. Ponadto, z punktu widzenia RODO i AI Act, czy DORA każde nieuprawnione ujawnienie danych wykorzystywanych w systemach AI może prowadzić do poważnych konsekwencji prawnych i regulacyjnych[124].

W rezultacie skutki incydentów w środowisku bankowym zintegrowanym z AI należy postrzegać nie tylko jako typowe dla klasycznych systemów IT, ale również jako złożone, dynamiczne i potencjalnie trudniejsze do wykrycia i usunięcia. Tym samym wzrasta znaczenie proaktywnego zarządzania ryzykiem, monitorowania modeli w czasie rzeczywistym oraz zapewniania przejrzystości i audytowalności systemów opartych na sztucznej inteligencji[125].

4.2.2. AI a ataki na klientów banków

Wraz z postępującą digitalizacją sektora finansowego, klienci banków coraz częściej stają się celem złożonych kampanii oszustw, w których cyberprzestępcy wykorzystują możliwości sztucznej inteligencji (AI). Technologia ta, choć wspiera automatyzację i efektywność procesów bankowych, staje się również potężnym narzędziem wykorzystywanym do ataków ukierunkowanych na osoby fizyczne.

Jednym z najbardziej niepokojących zjawisk jest zastosowanie technologii deepfake w kontekście inżynierii społecznej. Przestępcy wykorzystują syntetyczne obrazy twarzy lub wygenerowany głos, aby podszywać się pod pracowników banków, członków rodziny lub zaufane instytucje. Celem jest wywołanie zaufania i skłonienie ofiary do wykonania przelewu, ujawnienia danych logowania czy potwierdzenia kodu SMS. Jak wskazują eksperci CSIRT KNF, technologia ta staje się coraz bardziej dostępna i trudna do wykrycia dla przeciętnego użytkownika[126].

W 2024 roku w jednej z azjatyckich instytucji finansowych doszło do zaawansowanego oszustwa z wykorzystaniem technologii deepfake, które zyskało rozgłos w branży prawniczej i bankowej. Pracownik działu finansowego banku otrzymał wideokonferencyjne wezwanie od osoby, którą rozpoznał jako dyrektora regionalnego. Spotkanie odbyło się za pośrednictwem platformy komunikacyjnej, a uczestnicy, których twarze i głosy były realistyczne, przekazali pilne dyspozycje dotyczące przelewu środków na zagraniczne konta. Rozmowa zawierała szczegóły znane wyłącznie wewnętrznym pracownikom, co dodatkowo uwiarygodniło polecenie.

Rys. 7. Działania atakujących wykorzystujących AI dla deepfake.

Źródło: *Opracowanie własne.*

124 ENISA, *Multilayer Framework for Good Cybersecurity Practices for AI*, dz. cyt., s. 22–25; J. Surma, *Wprowadzenie do ataków na systemy uczenia maszynowego* [w:] A. Szczęsna, M. Stachoń (red), dz. cyt., s. 38–41.
125 ENISA, *Multilayer Framework for Good Cybersecurity Practices for AI*, dz. cyt., s. 22–27; J. Surma, *Wprowadzenie do ataków na systemy uczenia maszynowego* [w:] A. Szczęsna, M. Stachoń (red), dz. cyt., s. 42–45.
126 G. Gedye, *AI Voice Cloning: Do These 6 Companies Do Enough to Prevent Misuse?*, Consumer Reports 2025, 4 i n.

Jak się później okazało, nagranie wideo i głos przełożonego zostały wygenerowane za pomocą sztucznej inteligencji — był to deepfake, przygotowany przez przestępców z wykorzystaniem publicznie dostępnych materiałów wideo i informacji wykradzionych z wcześniejszych wycieków danych. Pracownik, nieświadomy oszustwa, zatwierdził przelew znacznej kwoty, co doprowadziło do realnej straty finansowej dla banku.

Ten przypadek doskonale ilustruje, jak AI może zostać wykorzystana do podszywania się pod osoby zaufane i przełamywania mechanizmów kontroli organizacyjnej, bazujących na relacjach osobistych i zaufaniu. W tym przypadku przestępcy nie użyli złośliwego oprogramowania, lecz manipulowali percepcją ofiary, wykorzystując zaawansowane techniki audiowizualnej symulacji.

Sztuczna inteligencja, a w szczególności technologie klonowania głosu, otwierają nowe, niepokojące możliwości dla cyberprzestępców. Zaledwie kilkusekundowy fragment nagrania audio wystarczy, by wygenerować realistyczną kopię głosu konkretnej osoby — członka rodziny, przełożonego, pracownika banku lub polityka. Ta technika, znana jako AI voice cloning, staje się jednym z najgroźniejszych narzędzi wykorzystywanych w tzw. vishingu (voice phishing) i oszustwach opartych na inżynierii społecznej[127].

Atakujący pozyskuje nagranie głosu ofiary (np. z mediów społecznościowych, nagrań konferencji lub rozmów online), a następnie wprowadza je do narzędzi takich jak ElevenLabs, PlayHT czy Parrot AI. Wygenerowany syntetyczny głos pozwala[128]:

- zadzwonić do ofiary podszywając się pod zaufaną osobę i wyłudzić dane lub środki,
- ominąć zabezpieczenia biometryczne (voice ID) w systemach bankowych – co wykazały m.in. testy Vice i BBC, gdzie dziennikarze złamali zabezpieczenia Santander, Halifax i innych banków,
- wywołać emocjonalną presję (np. symulacja płaczu dziecka lub paniki wnuczka),

Zastosowanie voice clone prowadzi do[129]:

- bezpośrednich strat finansowych (np. przelewy autoryzowane telefonicznie),
- przełamywania dwuskładnikowego uwierzytelnienia, jeśli jednym z czynników jest biometria głosu,
- wykorzystania danych wrażliwych (np. danych z bankowości elektronicznej, kodów SMS),
- traumatycznych skutków psychologicznych — wiele ofiar w badaniu amerykańskiego Consumer Report opisało uczucie „bycia zdradzonym przez własne zmysły", bo głos brzmiał identycznie jak bliskiej osoby.

Technika klonowania głosu uderza nie tylko w klientów, ale i w same instytucje[130]:

- utrata reputacji – banki postrzegane jako podatne na oszustwa mogą stracić zaufanie,
- nadużycia tożsamości w komunikacji z infolinią lub w zdalnej obsłudze klienta,
- przechodzenie przez procedury KYC z wykorzystaniem głosów „klientów", co otwiera drogę do prania pieniędzy,
- konieczność przeglądu stosowanych technologii biometrycznych, które – jak się okazuje – mogą być zbyt łatwe do obejścia.

W samym 2023 roku Federalna Komisja Handlu USA odnotowała ponad 850 tysięcy przypadków oszustw z podszyciem się, konsekwencją czego było ponad 2,7 miliarda USD strat finansowych. Co istotne, znaczna część z nich wiązała się z telefonicznym kontaktem i rosnącym wykorzystaniem narzędzi AI do klonowania głosu[131].

Z punktu widzenia klientów banków i pracowników działów obsługi, jest to szczególnie niebezpieczne zjawisko[132]. Coraz częściej mogą być oni celem oszustw, w których syntetyczne głosy i twarze będą wykorzystywane do:

- podszywania się pod infolinię bankową,
- kontaktowania się w sprawie „pilnych" operacji,

127 Tamże.
128 Tamże.
129 Tamże.
130 Tamże.
131 Tamże.
132 Zagrożenie to zostało również sklasyfikowane przez CETaS jako jedno z najbardziej zaawansowanych zastosowań AI w tzw. poważnych przestępstwach internetowych (ang. serious online crime). Technologia ta napędza nowe formy oszustw, które są trudne do wykrycia klasycznymi narzędziami, a jednocześnie bardzo skuteczne emocjonalnie. J. Burton, A. Janjeva, S. Moseley, A. Moseley, dz. cyt., s. 4 i n.

- autoryzowania przelewów na podstawie „wideoweryfikacji".

Zagrożenie to nabiera szczególnego znaczenia w dobie automatyzacji i digitalizacji bankowości, gdzie procesy weryfikacyjne coraz częściej opierają się na zdalnym kontakcie. Technologia deepfake, jeśli nie zostanie rozpoznana, może ominąć nawet wyrafinowane systemy uwierzytelniania, bazujące na rozpoznawaniu twarzy czy głosu[133].

Równolegle, sztuczna inteligencja wspiera rozwój zautomatyzowanych ataków phishingowych tekstowych. Generatywne modele językowe (np. GPT) umożliwiają tworzenie wysoce przekonujących, kontekstowych wiadomości e-mail, SMS-ów lub powiadomień, często dostosowanych do konkretnego klienta na podstawie jego cyfrowego śladu. Taka personalizacja zwiększa skuteczność ataków i obniża czujność odbiorcy. ENISA raportuje, że spoofing bankowych stron internetowych oraz podszywanie się pod infolinie bankowe (tzw. vishing) to jedne z głównych metod wykorzystywanych przeciwko klientom[134].

AI znajduje także zastosowanie w analizie danych z mediów społecznościowych, co umożliwia cyberprzestępcom segmentację ofiar i dostosowanie ataków do ich sytuacji życiowej, zainteresowań czy języka. Automatyzacja identyfikacji podatnych osób staje się elementem kampanii określanych jako „spear phishing na masową skalę"[135].

Najlepszym przykładem takich zaawansowanych działań jest ewolucja „przekrętów miłosnych". Romance fraud, czyli oszustwa emocjonalne oparte na zbudowaniu relacji uczuciowej, stały się w ostatnich latach jednym z najgroźniejszych typów przestępstw finansowych, szczególnie w połączeniu z technologiami AI. Przestępcy, korzystając z modeli językowych, generatorów deepfake'ów i analiz psychometrycznych, potrafią prowadzić wysoce spersonalizowane kampanie oszustw emocjonalno-finansowych[136].

AI wykorzystywana jest do[137]:

- automatycznego tworzenia tożsamości i profili w mediach społecznościowych, bazujących na fałszywych zdjęciach, historiach życiowych i stanowiskach (np. „inżynier na platformie wiertniczej" czy „amerykański lekarz na misji"),
- prowadzenia rozmów z ofiarą w czasie rzeczywistym przez boty wyposażone w zdolność analizy reakcji emocjonalnych,
- generowania fałszywych sytuacji kryzysowych, np. nagłych problemów zdrowotnych, konieczności opłacenia transportu lub uregulowania cła, w celu wymuszenia przelewów bankowych lub przesłania danych karty.

Zjawisko to coraz częściej łączy się z tzw. pig butchering scams, czyli długofalowym „tuczeniem" ofiary przez symulowaną relację, zakończonym przekierowaniem na fałszywą platformę inwestycyjną. Te fałszywe strony, stworzone z wykorzystaniem AI, symulują działające giełdy lub platformy kryptowalutowe, w których ofiara stopniowo inwestuje coraz większe środki, nieświadoma, że środki te są transferowane do przestępców[138].

Z punktu widzenia sektora bankowego, zagrożenie to ma wymiar zarówno indywidualny (utrata środków przez klientów), jak i instytucjonalny (utrata reputacji, wykorzystywanie kont bankowych do prania pieniędzy, obchodzenie KYC przy pomocy syntetycznych tożsamości generowanych przez AI)[139].

Nie można pominąć także zagrożenia związanego z kradzieżą tożsamości i fałszowaniem dokumentów. Algorytmy generatywne są wykorzystywane do tworzenia fałszywych zdjęć dokumentów, podpisów czy nawet wideo, co pozwala na obchodzenie procedur KYC (know your customer) w bankowości zdalnej[140].

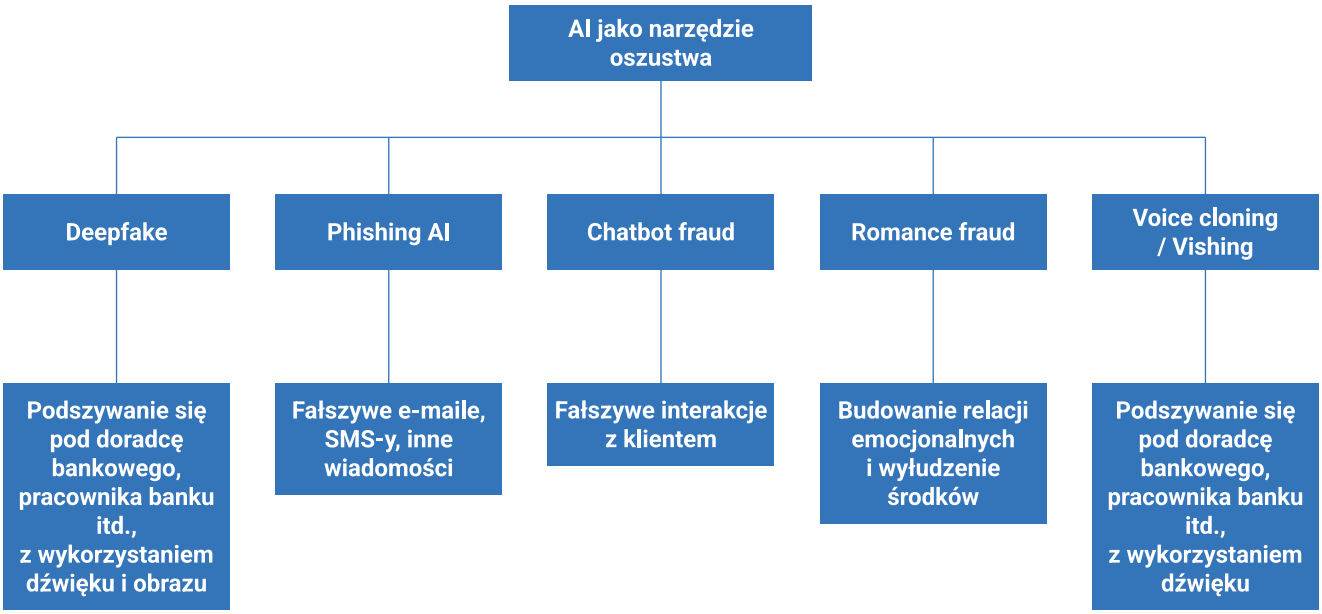
Konsekwencje takich działań są poważne: od utraty środków finansowych, przez trwałe naruszenie prywatności, aż po problemy prawne związane z wykorzystaniem tożsamości ofiary w dalszych przestępstwach. Co istotne, ofiary często nie są świadome,

133 G. Gedye, dz. cyt., s. 6–9.
134 ENISA, Threat Landscape: Finance Sector, dz. cyt., s. 9–11; J. Burton, A. Janjeva, S. Moseley, A. Moseley, dz. cyt., s. 4 i n.
135 K. Zieliński, A. Ślusarek, dz. cyt., s. 81–83.
136 S. Moseley, dz. Cyt., s. 3 i n.
137 Tamże.
138 Tamże.
139 Tamże.
140 J. Burton, A. Janjeva, S. Moseley, A. Moseley, dz. cyt., s. 18–21.

że padły celem ataku wspomaganego AI, co opóźnia reakcję i utrudnia śledztwo[141].

W tym kontekście kluczowe znaczenie ma budowanie świadomości klientów banków, edukacja w zakresie nowoczesnych metod cyberoszustw (rys. 8) oraz wdrażanie mechanizmów uwierzytelniania, które są odporne na manipulację generowaną komputerowo – np. uwierzytelnianie biometryczne z kontrolą żywotności (liveness detection) czy analiza behawioralna[142].

Rys. 8. Omówione oszustwa, które ewoluują, dzięki AI.



Źródło: *Opracowanie własne.*

Jak podsumowuje to opinia wydana przez EBA (EAB/Op/2025/10) w punkcie 12 i 13: „Ekspansja cyberprzestępczości i oszustw (…) nadal przewyższa możliwości obronne sektora. (…) Przeciwdziałanie tym zagrożeniom będzie wymagało zaawansowanych rozwiązań technicznych i specjalistycznej wiedzy."[143].

4.3. Błędy systemowe i wpływ AI na stabilność usług płatniczych

Błędy systemowe oraz wykorzystanie AI istotnie wpływają na stabilność usług płatniczych. Z analiz zawartych w raporcie „Cyberbezpieczeństwo AI. AI w cyberbezpieczeństwie" można wyodrębnić trzy kluczowe wnioski:

1. Wysoka zależność usług finansowych od infrastruktury cyfrowej[144]

   Usługi płatnicze w dużej mierze przeniesione zostały do domeny cyfrowej, co sprawia, że każda awaria systemowa, błąd w konfiguracji, utrata łączności czy atak DDoS może prowadzić do destabilizacji systemu i wstrzymania operacji finansowych. Analogowe alternatywy funkcjonują głównie w ramach planów ciągłości działania (BCP), co zwiększa ryzyko w przypadku ich niedostępności.

2. Omówione wcześniej złośliwe wykorzystanie AI przez cyberprzestępców[145]

   Systemy płatnicze stają się celem zaawansowanych ataków z użyciem AI, m.in. poprzez techniki deepfake służące do podszywania się pod klientów lub pracowników w celu wyłudzenia środków. Użycie zmanipulowanych treści głosowych lub wideo w systemach autoryzacji biometrycznej może prowadzić do nieautoryzowanych transakcji i strat finansowych.

141 J. Burton, A. Janjeva, S. Moseley, A. Moseley, dz. cyt., s. 21–24; S. Moseley, dz. Cyt., s. 7–10.
142 S. Moseley, dz. Cyt., s. 11–13.
143 EBA, *Opinion of the European Banking Authority on money laundering and terrorist financing risks affecting the EU's financial sector*, EBA/Op/2025/10, 2025.
144 K. Silicki, dz. cyt., s. 10 i n.
145 K. Zieliński, A. Ślusarek, dz. cyt., s. 82–99.

3. Omówione wcześniej ryzyko błędów modeli AI i ich wpływ na zaufanie[146]

   Błędy lub awarie w algorytmach AI (np. przeuczenie modelu, manipulacja danymi treningowymi, błędne etykiety) mogą skutkować nieprawidłowym działaniem systemów rekomendacyjnych i decyzyjnych, np. w kontekście przyznawania kredytów, wykrywania oszustw czy zarządzania ryzykiem. Może to wpłynąć nie tylko na stabilność techniczną, ale i na utratę zaufania klientów.

Podsumowując, aby zachować stabilność usług płatniczych w erze AI, konieczne jest wdrażanie środków zaradczych w zakresie technicznego bezpieczeństwa systemów, a także rozwijanie odporności organizacyjnej i zarządczej na zagrożenia płynące z wykorzystania sztucznej inteligencji.

4.4. Kwestie „odpowiedzialności"

Odpowiedzialność stanowi jeden z kluczowych elementów zaufania do AI. W ujęciu etycznym, technicznym i regulacyjnym odpowiedzialność odnosi się do konieczności przypisania jasnych ról, obowiązków i możliwości kontroli nad działaniami systemu AI w całym jego cyklu życia. Jest ona nieodzowna do zapewnienia zgodności z prawami podstawowymi, jak również do zapobiegania szkodom wynikającym z autonomicznych decyzji systemów. W „Wytycznych w zakresie etyki dotyczącej godnej zaufania sztucznej inteligencji" Komisji Europejskiej, odpowiedzialność (ang. accountability) została wskazana jako jeden z siedmiu kluczowych wymogów godnej zaufania SI. Zaznacza się, że: „Należy zapewnić opracowywanie, wdrażanie i wykorzystywanie systemów SI zgodnie z wymogami dotyczącymi godnej zaufania sztucznej inteligencji: […] 7) odpowiedzialność". Wytyczne podkreślają konieczność dokumentowania decyzji, ocen kompromisów między wymogami oraz angażowania interesariuszy w cały cykl życia AI[147]. Ponadto raport „Ahead of the Curve" analizujący zastosowanie AI Act do agentów AI wskazuje, że: „Zarządzanie agentami SI ilustruje problem 'wielu rąk' (many hands problem), w którym odpowiedzialność jest niejasna ze względu na rozproszony łańcuch wartości". W odpowiedzi na to wyzwanie AI Act przypisuje obowiązki trzem głównym kategoriom podmiotów:

- dostawcom modeli ogólnego przeznaczenia (GPAI),
- dostawcom systemów,
- wdrażającym systemy.

Każda z tych grup ma własne obowiązki w zakresie oceny ryzyka, przejrzystości, nadzoru technicznego i odpowiedzialności za skutki działania systemów AI[148]. Z kolei analizując holenderski przykład tj. AI Impact Assessment (AIIA) opracowanego przez holenderskie Ministerstwo Infrastruktury i Gospodarki Wodnej, cały rozdział 7 poświęcono odpowiedzialności. Uwzględnia on takie zagadnienia jak:

- przejrzystość wobec użytkownika (7.1),
- komunikacja z interesariuszami (7.2),
- weryfikowalność decyzji (7.3),
- archiwizacja działań (7.4).

AIIA wymaga, by zespół projektowy jednoznacznie określił, kto ponosi odpowiedzialność za decyzje wspierane lub podejmowane przez AI[149]. Nie inaczej da się zaobserwować w kazusie niemieckim, gdzie Niemiecki Federalny Urząd ds. Bezpieczeństwa Informacji opracował katalog testowy dla AI w finansach, który zawiera kryteria[150]:

- GOV-02: ustanowienie systemu zarządzania SI i przeprowadzanie audytów,
- HO-04: przypisanie odpowiedzialności etycznej i prawnej za cały cykl życia SI.

To podejście wymaga m.in. wskazania osób odpowiedzialnych, stworzenia katalogu modeli AI i zapewnienia kompetencji zespołu nadzorczego.

Norma ISO/IEC 42001:2023, dotycząca systemów zarządzania SI, traktuje odpowiedzialność jako kluczowy element zarządzania ryzykiem i zgodności z regulacjami. Wdrażając ISO 42001, organizacje mogą wykazać zaangażowanie w etyczne praktyki SI i zdobyć zaufanie interesariuszy. Zarządzanie odpowiedzialnością obejmuje polityki wewnętrzne, audyty, przypisanie ról i mechanizmy odpowiedzi

146 M. Krzyszton, *Weryfikacja wiarygodności systemów w erze uczenia maszynowego* [w;] A. Szczęsna, M. Stachoń dz. cyt., s. 45–58.
147 Grupa ekspertów wysokiego szczebla ds. SI, *Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji*, …, s. 3.
148 A. Oueslati, R. States-Polet, …, s. 25–28.
149 Ministry of Infrastructure and Water Management, …, s. 26–28.
150 Federal Office for Information Security, …, s. 106–105 oraz 137–138.

na incydenty[151]. W opracowaniu „AI Explainability in Practice" opracowanym przez The Alan Turing Institute odpowiedzialność została ujęta jako jedna z czterech głównych zasad (maxims) wyjaśnialności AI: „Zasada 2: Bądź odpowiedzialny – zdefiniuj, kto odpowiada za projekt, wdrożenie i wyniki działania systemu SI". Odpowiedzialność łączy się tu bezpośrednio z wymogiem wyjaśniania decyzji i rozliczalności zespołów projektowych[152].

Autorzy raportu po przeanalizowaniu zebranych dokumentów, których treść została przytoczona powyżej, uważają, że można wyróżnić trzy główne wymiary odpowiedzialności w kontekście AI:

1. Strukturalna – przypisanie ról i odpowiedzialności na poziomie organizacyjnym (np. zgodnie z AI Act, ISO/IEC 42001:2023).

2. Procesowa – mechanizmy zapewniające rozliczalność (audyt, dokumentacja, monitoring).

3. Etyczna i społeczna – odpowiedzialność za skutki społeczne, wpływ na prawa podstawowe i potrzebę zachowania ludzkiego nadzoru[153].

Rozpoznanie i egzekwowanie odpowiedzialności jest warunkiem koniecznym do budowania zaufania do AI i zgodnego z prawem jej wykorzystania.

151 S. Musch, M. Borrelli, C. Kerrigan, J. Bennison, …, s. 5–6.

152 D. Leslie,, C. Rincón, M. Briggs, A. Perini, S. Jayadeva, A. Borda, C. SJ. Burr Bennett, M. Aitken, S. Mahomed, J. Wong, M. Waller, C. Fischer, …, s. 18–19.

153 Inny raport przygotowany na potrzeby WIB, na temat postrzegania AI przez polskich konsumentów pokazuje potrzebę uwzględnienia społecznych obaw związanych z bezpieczeństwem danych, etyką decyzji AI i utratą pracy. K. Sekścińska, A…, s. 7–8, 36–38, 71 oraz 74–76.

5. AI Act i jego wpływ na sektor płatniczy

Wraz ze wskazanym dynamicznym rozwojem systemów sztucznej inteligencji, europejski ustawodawca podjął próbę kompleksowej regulacji tego obszaru, czego efektem jest AI Act – pierwsze na świecie ogólne rozporządzenie dotyczące systemów sztucznej inteligencji. AI Act ustanawia ramy prawne oparte na analizie ryzyka, obejmując pełen cykl życia systemów AI – od projektowania, przez wdrożenie, aż po eksploatację – oraz nakładając konkretne obowiązki na dostawców, użytkowników, importerów i dystrybutorów tych technologii. Choć akt ten ma charakter horyzontalny, jego skutki szczególnie wyraźnie zaznaczają się w sektorze finansowym, w tym w usługach płatniczych. AI Act bezpośrednio wpływa na te praktyki, ponieważ systemy służące do oceny zdolności kredytowej lub ustalania punktacji kredytowej konsumentów zostały uznane za systemy wysokiego ryzyka (high-risk AI systems) i podlegają rygorystycznym wymogom zgodności. AI Act odwołuje się też do aktów legislacyjnych, czy też narzędzi prawno technologicznych wynikających z prawa finansowego (rys. 9).

Przykładowo, jednym z kluczowych zastosowań AI w sektorze płatniczym jest automatyzacja oceny zdolności kredytowej konsumentów, oparta na algorytmach przetwarzających dane historyczne, dane behawioralne czy ślady cyfrowe[154]. Fintechy i instytucje finansowe coraz częściej wykorzystują systemy oparte na AI do analizowania tzw. digital footprint, co może zwiększać dostęp do finansowania, ale jednocześnie rodzi ryzyko dyskryminacji i obniżenia transparentności decyzji[155]. Takie niestandardowe dane to np.[156]:

- domeny e-mail (np. Hotmail, Gmail),
- czas dnia, w którym użytkownik wykonuje zakupy,
- długość tekstu, który wpisuje w formularzu,
- dane z telefonu komórkowego, w tym żywotność baterii i logi połączeń.

AI Act klasyfikuje takie systemy jako „wysokiego ryzyka", co oznacza obowiązek wdrożenia rygorystycznych procedur oceny ryzyka, testowania jakości danych oraz zapewnienia wyjaśnialności decyzji podejmowanych przez algorytmy.

Kolejnym istotnym obszarem jest wykrywanie oszustw finansowych i przeciwdziałanie praniu pieniędzy (AML). W tym kontekście AI pozwala na analizowanie wzorców transakcyjnych w czasie rzeczywistym, co zwiększa skuteczność identyfikacji nadużyć. AI Act, choć wspiera rozwój takich rozwiązań, wymaga od ich dostawców spełnienia kryteriów bezpieczeństwa, dokładności i odporności na błędy systemowe – w tym dokumentowania procesu uczenia modeli oraz zapewnienia ludzkiego nadzoru nad ich działaniem.

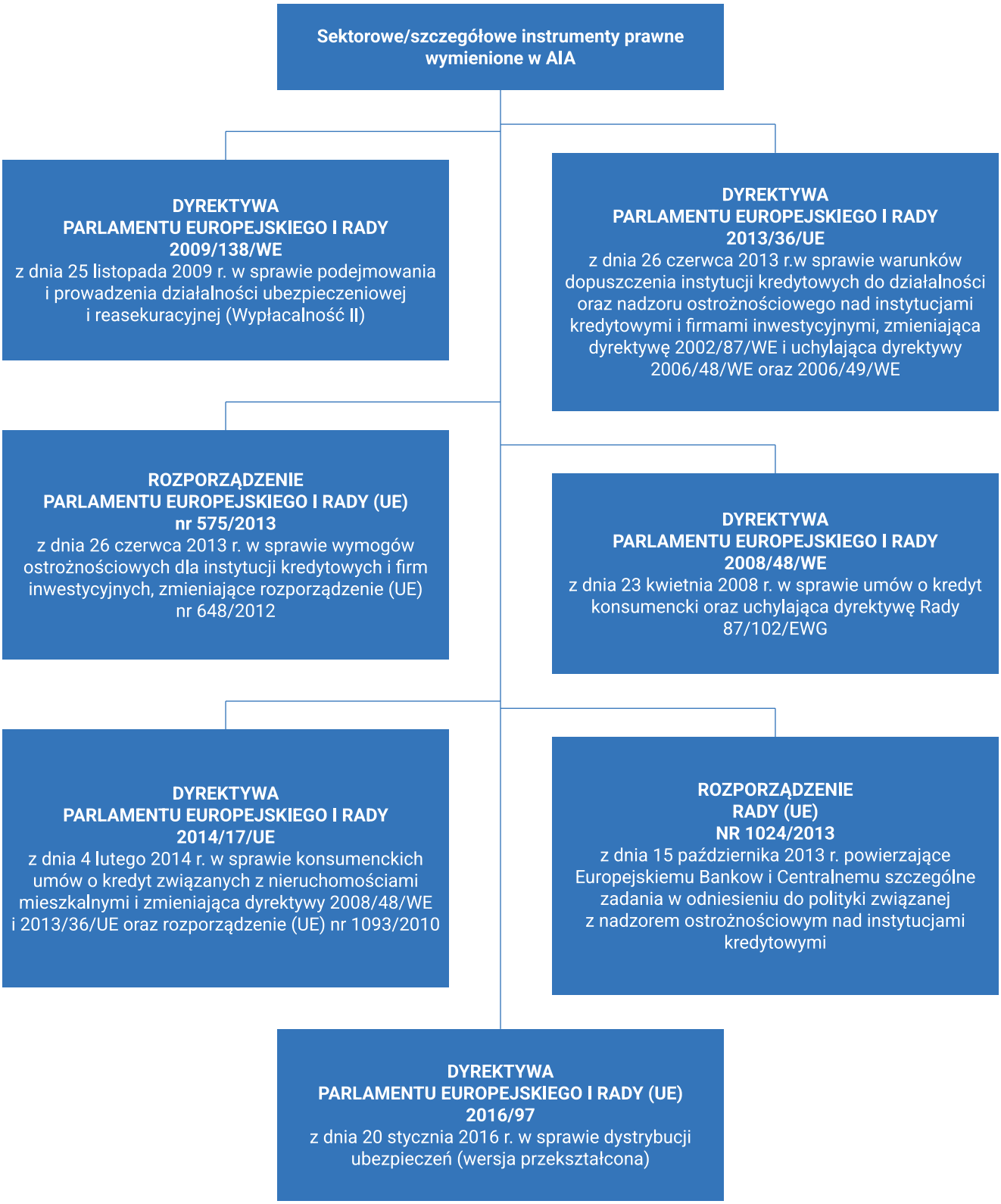
W sektorze płatniczym popularne stają się też chatboty i wirtualni asystenci wykorzystywani do obsługi klienta, w tym do pomocy w inicjowaniu przelewów, sprawdzaniu salda czy rozwiązywaniu problemów technicznych. Choć AI Act nie uznaje wszystkich takich systemów za wysokiego ryzyka, to w przypadku, gdy wirtualny agent wpływa na decyzje finansowe użytkownika (np. rekomendując produkt kredytowy), może już podlegać surowszym regulacjom, w tym obowiązkom informacyjnym i zapewnieniu przejrzystości działania algorytmu.

154 Jak wskazuje K. Szczudlik: „Banki, inne instytucje ustawowo upoważnione do udzielania kredytów, instytucje pożyczkowe oraz podmioty, o których mowa w art. 59d ustawy z dnia 12 maja 2011 r. o kredycie konsumenckim, a także instytucje utworzone na podstawie art. 105 ust. 4, mogą w celu oceny zdolności kredytowej i analizy ryzyka kredytowego podejmować decyzje, opierając się wyłącznie na zautomatyzowanym przetwarzaniu, w tym profilowaniu, danych osobowych – również stanowiących tajemnicę bankową – pod warunkiem zapewnienia osobie, której dotyczy decyzja podejmowana w sposób zautomatyzowany, prawa do otrzymania stosownych wyjaśnień co do podstaw podjętej decyzji, do uzyskania interwencji ludzkiej w celu podjęcia ponownej decyzji oraz do wyrażenia własnego stanowiska.". Zob. K. Szczudlik, *DORA a sztuczna inteligencja*, Wolters Kluwer 2024, el.

155 Zob. D. Szostek, M. Nowakowski, G. Bar, R. Prabucki, dz. cyt, s. 25–27, 47–49, 51–52.

156 Jak wskazuje Wolfgang Spiess-Knafl, niektóre instytucje finansowe już dziś wykorzystują digital footprint – przykłady firm takich jak Klarna, Sesame Credit, ZestFinance, Branch, czy KrediTech, pokazują, że są podmioty, które wykorzystują alternatywne dane do predykcji zachowań kredytowych. W. Spiess-Knafl, dz. cyt, s. 66–70.

Rys. 9. Odwołania w AI Act do unijnego prawa finansowego. Źródło: opracowanie własne.



Źródło: *Opracowanie własne.*

Wreszcie, dynamiczne ustalanie cen i opłat transakcyjnych, bazujące na zachowaniu użytkownika, czasie i lokalizacji – stosowane w aplikacjach mobilnych czy portfelach cyfrowych – budzi kontrowersje w zakresie personalizowanego ustalania warunków, które mogą prowadzić do naruszenia zasad równego traktowania. AI Act wymaga, by takie systemy były poddane ocenie pod kątem ryzyka dyskryminacji oraz zawierały środki zaradcze w przypadku wykrycia nieprawidłowości.

Z perspektywy sektora płatniczego oznacza to konieczność wdrożenia nie tylko odpowiednich procedur zarządzania ryzykiem, ale też systemów nadzoru ludzkiego, mechanizmów wyjaśnialności algorytmów oraz prowadzenia dokumentacji zgodnej z rozporządzeniem. W szczególności wymagane będzie zapewnienie jakości danych wykorzystywanych do trenowania algorytmów scoringowych, ich audytowalność, a także analiza i minimalizacja ryzyka dyskryminacji

Należy też zaznaczyć, że AI Act nie stosuje się do systemów AI udostępnianych na podstawie bezpłatnych licencji otwartego oprogramowania, chyba, że systemy te są wprowadzane do obrotu lub oddawane do użytku jako system AI wysokiego ryzyka lub jako system AI objęty art. 5 lub art. 50 AI Act.

Wprowadzenie AI Act istotnie zmienia sposób projektowania i wdrażania systemów AI w płatnościach – wymuszając większą kontrolę nad ich zgodnością, przejrzystością i wpływem na prawa podstawowe użytkowników. Jak wskazują P. Hense i T. Mustać, skuteczna adaptacja do nowych ram prawnych wymaga nie tylko analizy technicznej ryzyk, ale także zintegrowanego podejścia do compliance i zarządzania cyklem życia systemu AI – od momentu projektowania, przez dokumentację, aż po nadzór ex post w środowisku korporacyjnym[157].

5.1. Regulacje dotyczące systemów AI wysokiego ryzyka

Zgodnie z AI Act, systemy sztucznej inteligencji (AI) wysokiego ryzyka są objęte szczegółowymi regulacjami prawnymi ze względu na ich potencjalny wpływ na zdrowie, bezpieczeństwo oraz prawa podstawowe osób fizycznych.

System AI uznawany jest za wysokiego ryzyka, jeżeli[158]:

1. stanowi komponent bezpieczeństwa produktu podlegającego regulacjom wymienionym w załączniku I (np. urządzenia medyczne, zabawki, maszyny) lub

2. jest systemem autonomicznym, stosowanym w jednym z ośmiu obszarów wymienionych w załączniku III, takich jak:
   - biometryka (np. zdalna identyfikacja biometryczna),
   - infrastruktura krytyczna (np. zarządzanie ruchem drogowym),
   - edukacja i szkolenia zawodowe,
   - zatrudnienie i zarządzanie pracownikami,
   - dostęp do usług prywatnych i publicznych,
   - egzekwowanie prawa,
   - migracja i kontrola graniczna,
   - wymiar sprawiedliwości i procesy demokratyczne

Systemy te muszą spełniać tzw. „istotne wymagania" (art. 8–15 AI Act), w tym[159]:

- system zarządzania ryzykiem przez cały cykl życia systemu (art. 9),
- wysoką jakość danych i zarządzanie danymi treningowymi, testowymi i walidacyjnymi (art. 10),
- dokumentacja techniczna i rejestrowanie zdarzeń (art. 11 i 12),
- transparentność — system musi informować użytkowników o swoim przeznaczeniu i ograniczeniach (art. 13),
- nadzór ludzki — system musi umożliwiać efektywny nadzór przez człowieka (art. 14),
- dokładność, solidność i cyberbezpieczeństwo (art. 15)

Dostawcy systemów sztucznej inteligencji wysokiego ryzyka zobowiązani są do prowadzenia obszernej dokumentacji technicznej, która musi zostać przygotowana przed wprowadzeniem systemu do obrotu. Dokumentacja ta powinna obejmować m.in. opis systemu, jego przeznaczenie, zastosowane algorytmy, dane użyte do trenowania, środki kontroli ryzyka, a także wyniki przeprowadzonych testów i ocen zgodności. Obowiązek ten wynika wprost z załącznika IV do AI Act oraz art. 11 i 17 rozporządzenia. Ponadto, dostawcy muszą prowadzić rejestry automatycznie generowanych zdarzeń przez systemy AI oraz zapewnić, że systemy te wyposażone są w funkcje umożliwiające monitorowanie ich działania przez człowieka. Systemy te muszą być również stale

157 P. Hense, T. Mustać, dz. Cyt., s. 10–12, 32–36.
158 Zob. AI Act, art. 6 ust. 1–2 oraz załączniki I i III; N. A. Smuha, K. Yeung, *The European Union's AI Act: Beyond Motherhood and Apple Pie?* [w:] N. A. Smuha (red.), dz. Cyt., s. 228–260.
159 Zob. AI Act, art. 29–30 oraz art. 61; N. A. Smuha, K. Yeung, dz. Cyt., s. 242–246.

monitorowane po wprowadzeniu na rynek zgodnie z opracowanym planem nadzoru – ma to na celu zapewnienie ich zgodności z wymogami prawnymi i wykrywanie ewentualnych zagrożeń dla praw podstawowych użytkowników[160].

Większość systemów AI wysokiego ryzyka może przejść ocenę zgodności w trybie wewnętrznym (self-assessment), poza niektórymi przypadkami dotyczącymi systemów biometrycznych, które wymagają oceny przez notyfikowane jednostki certyfikujące[161]. Ocena ta obejmuje[162]:

- wdrożenie systemu zarządzania jakością (QMS),
- przygotowanie dokumentacji technicznej zgodnie z załącznikiem IV,
- monitorowanie ryzyka po wprowadzeniu systemu na rynek

Dostawcy i użytkownicy systemów wysokiego ryzyka mają obowiązek[163]:

- zapewnienia aktualizacji dokumentacji,
- prowadzenia systemu monitorowania po wdrożeniu,
- przeprowadzania oceny wpływu na prawa podstawowe (dla podmiotów publicznych i prywatnych świadczących usługi publiczne)

Wysokie ryzyko wiąże się z koniecznością szczegółowego testowania (art. 9 ust. 6–8) oraz utrzymania wysokiej jakości danych (art. 10), w tym:

- analizy biasów,
- kompletności i reprezentatywności zbiorów danych,
- ochrony danych wrażliwych przy spełnieniu warunków prawnych.

Ponadto, systemy AI wysokiego ryzyka muszą być odporne na ataki, zgodnie z art. 15 AI Act, dostawcy systemów wysokiego ryzyka muszą zapewnić ich[164]:

- odporność na manipulacje danymi wejściowymi (np. adversarial examples),
- odporność na złośliwe ingerencje w modele (np. model poisoning),
- zdolność do odparcia prób wydobycia danych uczących (membership inference, privacy leakage).

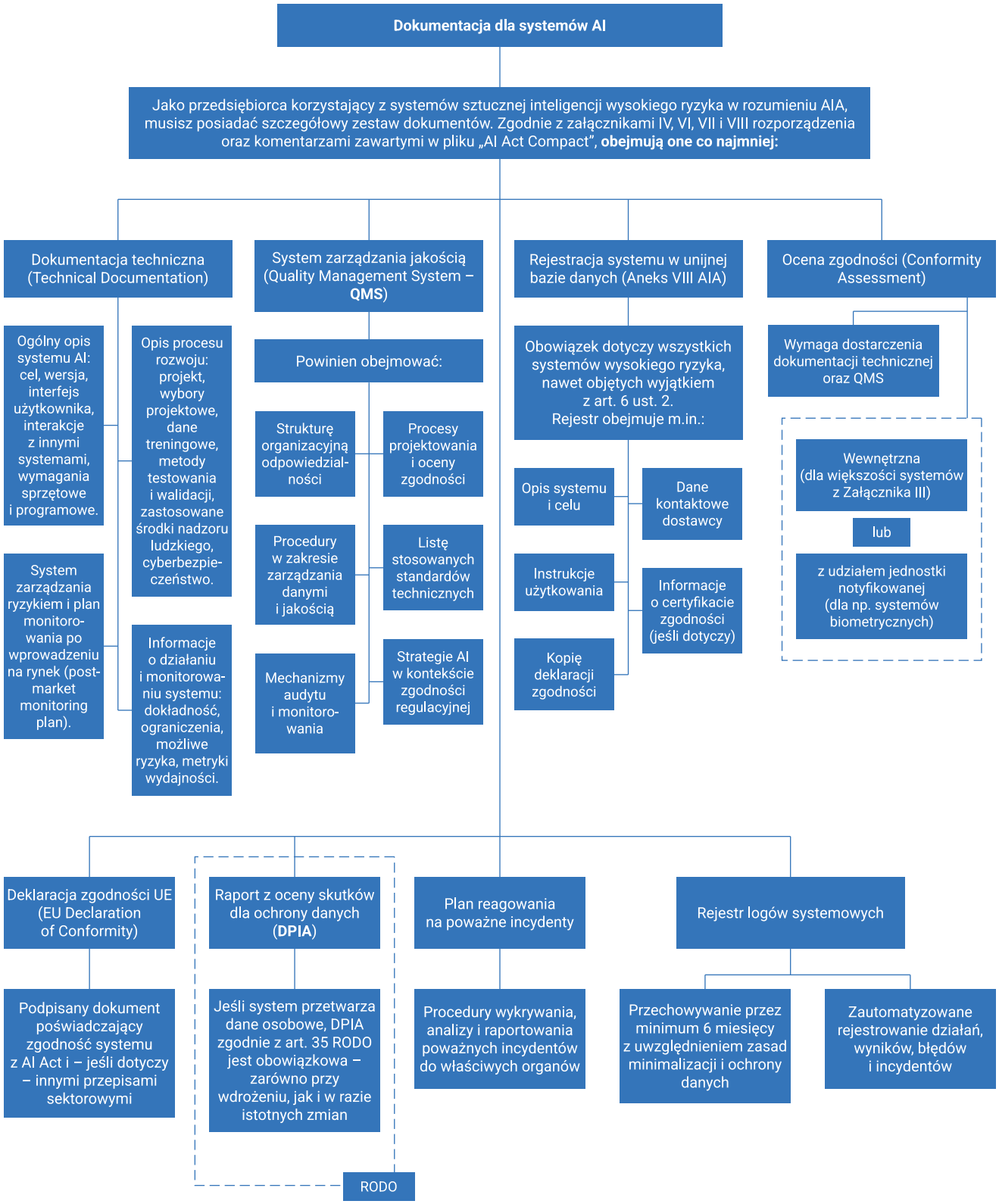
Cyberbezpieczeństwo w AI Act jest rozumiane szerzej niż w tradycyjnym podejściu CIA (confidentiality, integrity, availability) i uwzględnia specyfikę systemów uczących się. Oznacza to przesunięcie akcentu z klasycznych zabezpieczeń infrastrukturalnych na ochronę przed atakami unikalnymi dla uczenia maszynowego i danych treningowych. AI Act wprost przewiduje obowiązek wdrażania środków zapobiegawczych i reagowania na takie zagrożenia. Generalnie istnieją cztery kategorie zagrożeń, które powinny być brane pod uwagę przy projektowaniu i ocenie systemów wysokiego ryzyka:

- manipulacja zachowaniem systemu (np. poprzez fałszywe dane wejściowe),
- utrata integralności lub nieprzewidywalne działanie modelu,
- naruszenia prywatności wynikające z metod trenowania modelu,
- eksfiltracja danych lub kodu poprzez interakcję z modelem.

Wskazuje się również, że tradycyjne podejścia do oceny ryzyka nie są wystarczające. Potrzebne są nowe metody testowania odporności modeli — zbliżone do tych rozwijanych przez MITRE ATLAS, ENISA, czy w ramach NIST AI RMF[165].

160 A. Szczęsna, *AI Act – wyczekiwana regulacja systemów sztucznej inteligencji wysokiego ryzyka* [w:] A. Szczęsna, M. Stachoń, dz. cyt., s. 142–143.
161 Zob. AI Act, art. 43 ust. 1–2 i 4 oraz załącznik VII; N. A. Smuha, K. Yeung, dz. Cyt., s. 240–241.
162 Zob. AI Act, art. 17–23 oraz załączniki IV i VII; por. P. Hense, T. Mustać, dz. Cyt., s. 72–74, gdzie szczegółowo omówiono obowiązki w zakresie systemu zarządzania jakością, dokumentacji technicznej oraz monitorowania po wprowadzeniu do obrotu.
163 Zob. *AI Act*, art. 29 (obowiązki dostawców), art. 30 (obowiązki użytkowników) oraz art. 61 (ocena wpływu na prawa podstawowe); N. A. Smuha, K. Yeung, „dz. cyt, s. 243–246.
164 P. Hense, T. Mustać, dz. Cyt., s. 94–105.
165 J. Surma, *Wprowadzenie do ataków na systemy uczenia maszynowego* [w:] A. Szczęsna, M. Stachoń (red), dz. cyt., s. 34–36; D. Szostek, P. Kasprowski, J. Kozak, A. Kapczyński, R. Prabucki, *Wyzwania i zagrożenia z zakresu cyberbezpieczeństwa podczas projektowania lub wykorzystywania AI* [w:] A. Szczęsna, M. Stachoń (red), dz. cyt., s. 28–33.

Rys. 10. Podstawowa dokumentacja, wymagana przy systemach AI wysokiego ryzyka.



Źródło: *R. Prabucki, Wykorzystanie AI w Compliance, Politechnika Śląska 2025.*

5.2. Wymogi w zakresie zgodności i audytowalności systemów AI

Zgodnie z AI Act, systemy sztucznej inteligencji wysokiego ryzyka podlegają szczegółowym wymogom w zakresie zgodności, które mają zapewnić ich bezpieczne, przejrzyste i zgodne z prawami podstawowymi funkcjonowanie. Kluczowym elementem w tym kontekście jest zdolność do zapewnienia audytowalności i rozliczalności tych systemów na każdym etapie ich cyklu życia.

Podstawą oceny zgodności jest zdefiniowany w art. 17 AI Act system zarządzania jakością, który dostawca systemu AI musi ustanowić, wdrożyć i utrzymywać. System ten obejmuje m.in. procedury zarządzania ryzykiem (art. 9), wymagania dotyczące jakości i bezpieczeństwa danych (art. 10), dokumentację techniczną (art. 18), procedury monitorowania po wprowadzeniu do obrotu (art. 61) oraz mechanizmy informowania organów nadzorczych (art. 73).

Jednostki notyfikowane, uprawnione przez państwa członkowskie do przeprowadzania oceny zgodności (art. 30–36 i załącznik VI), odgrywają kluczową rolę w procesie audytu. Weryfikują one, czy dostawca skutecznie wdrożył system zarządzania jakością i czy jest on stosowany w praktyce. W tym celu jednostki mają dostęp do dokumentacji technicznej każdego systemu objętego zakresem audytu, mogą przeprowadzać inspekcje w siedzibach projektowych i testowych, a także wykonywać dodatkowe testy systemów AI — nawet jeśli wcześniej uzyskały one certyfikat zgodności na podstawie dokumentacji.

Po przeprowadzeniu audytu jednostka notyfikowana sporządza sprawozdanie z audytu, w którym przedstawia ocenę funkcjonowania systemu zarządzania jakością, wskazuje ewentualne niezgodności i zaleca działania naprawcze. Procedura ta zapewnia mechanizm bieżącej kontroli oraz przyczynia się do budowy zaufania do systemów AI.

Wymóg audytowalności nie ogranicza się wyłącznie do aspektów formalnych. Odnosi się również do technicznej zdolności systemu AI do śledzenia, uzasadniania i dokumentowania procesu decyzyjnego (np. przy użyciu rejestrów, logów lub ścieżek wyjaśniających), co jest kluczowe z perspektywy przejrzystości, zgodności z prawem oraz możliwości wykazania zgodności z wymaganiami rozporządzenia w przypadku kontroli lub incydentu.

W praktyce zapewnienie zgodności i audytowalności wymaga nie tylko spełnienia wymagań technicznych, ale również opracowania i wdrożenia odpowiednich procedur organizacyjnych, dokumentacyjnych i kontrolnych. Takie podejście wzmacnia zasadę rozliczalności (accountability) i wspiera realizację głównego celu AI Act: zapewnienia, że systemy AI działają w sposób bezpieczny, godny zaufania i zgodny z wartościami Unii Europejskiej.

5.3. Wpływ braku jednoznacznej definicji AI na wdrożenia regulacyjne

Brak jednoznacznej i precyzyjnej definicji sztucznej inteligencji (AI) stanowi istotną barierę dla skutecznego wdrażania przepisów prawa w sektorze płatniczym, w którym automatyzacja procesów, analiza wzorców transakcyjnych i systemy wykrywania nadużyć odgrywają kluczową rolę. Choć unijne regulacje, w szczególności Akt o sztucznej inteligencji (AI Act), próbują objąć zakresem szerokie spektrum zastosowań AI, to jednak niejasności definicyjne prowadzą do poważnych trudności interpretacyjnych.

Zgodnie z art. 3 pkt 1 AI Act, systemem sztucznej inteligencji jest oprogramowanie opracowane z wykorzystaniem jednej lub większej liczby technik i podejść wymienionych w załączniku I, które może, w przypadku zastosowania, generować wyniki takie jak treści, przewidywania, zalecenia lub decyzje wpływające na środowisko. Definicja ta, oparta pierwotnie na deklaracji OECD, obejmuje wiele rozwiązań stosowanych na rynku płatniczym, takich jak systemy scoringowe, silniki decyzyjne czy mechanizmy przeciwdziałania praniu pieniędzy (AML) wykorzystujące reguły i proste algorytmy heurystyczne[166].

W rezultacie, nawet klasyczne narzędzia regułowe – stosowane w monitoringu transakcji – mogą być objęte wymogami AI Act, mimo że nie posiadają żadnych zdolności adaptacyjnych. Może to prowadzić do nieproporcjonalnych obowiązków zgodności, np. konieczności przeprowadzania oceny zgodności ex ante, czy notyfikowania systemów w rejestrze UE.

AI Act, m.in. w motywie 12 preambuły, stara się rozróżnić AI od tradycyjnych systemów informatycz-

166 OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, 2019. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449 (dostęp: 08.05.2025).

nych, sugerując, że system AI cechuje się autonomią i zdolnością do wyciągania wniosków z danych wejściowych. Jednak w praktyce różnica ta bywa trudna do uchwycenia – np. w przypadku systemów płatności, które automatycznie blokują podejrzane transakcje lub analizują wzorce zachowań konsumenckich na podstawie reguł i klasyfikatorów. Nie jest jednoznaczne czy tego typu rozwiązania są „AI" w rozumieniu rozporządzenia[167].

Instytucje płatnicze – w tym dostawcy usług płatniczych (PSP) i instytucje pieniądza elektronicznego – są zobowiązane do zachowania zgodności nie tylko z AI Act, ale także z przepisami AMLD, PSD2, DORA i CRD/CRR. Niejednoznaczność definicji AI skutkuje koniecznością stosowania konserwatywnego podejścia i często oznacza nadmierne obciążenia regulacyjne w odniesieniu do systemów, które nie generują istotnych ryzyk[168].

Przykładowo, wdrożenie nowego silnika transakcyjnego do dynamicznego ustalania limitów może wymagać – z ostrożności – pełnej procedury oceny zgodności z AI Act, co wpływa na czas wdrożenia oraz nakłady na dokumentację techniczną, testy i nadzór.

Organy nadzoru rynku finansowego, w tym europejskie (EBA) i krajowe (np. KNF), mogą różnie interpretować pojęcie „systemu AI". Brak jednoznaczności może prowadzić do tzw. forum shopping, czyli celowego lokowania operacji w jurysdykcjach bardziej liberalnie interpretujących definicję, co osłabia jednolitość rynku wewnętrznego i stwarza asymetrie konkurencyjne[169].

Zbyt szerokie lub niejasne zdefiniowanie AI może skutkować nadregulowaniem rozwiązań niskiego ryzyka – jak np. chatbotów obsługujących zapytania o saldo czy predykcyjnych mechanizmów personalizacji ofert – co z kolei może zniechęcać podmioty do wdrażania nowych technologii. Jednocześnie zbyt liberalne podejście do definicji może prowadzić do niedoregulowania rozwiązań rzeczywiście wpływających na decyzje kredytowe lub wykrywanie oszustw[170].

W literaturze podkreśla się potrzebę wprowadzenia bardziej operacyjnych definicji lub kryteriów funkcjonalnych, które lepiej oddzielą klasyczne oprogramowanie od rzeczywistej AI. W przypadku sektora płatniczego, pomocne mogłoby być np. przyjęcie wytycznych sektorowych opartych na ryzyku (risk-based guidance) lub wskazanie typowych kazusów podlegających lub wyłączonych spod AI Act.

167 Warto tutaj podkreślić, że sprawa ta jest na tyle kontrowersyjna, że dotknęła nawet pytanie prejudycjalne dot. polskiego systemu SLPS, tj. Systemu Losowego Przydziału Spraw (dla sędziów). Zob. TSUE oceni system losowania, DGP 2025.

168 A. Keller, C. M. Pereira, M. L. Pires, *The European Union's Approach to Artificial Intelligence and the Challenge of Financial Systemic Risk* [w:] H. S. Antunes, P. M. Freitas, A. L. Oliveira, C. M. Pereira, E. V. de Sequeira, L. B. Xavier (red.), *Multidisciplinary Perspectives on Artificial Intelligence and the Law, Vol. 58*, Springer 2024, s. 415–438; Zob. też P. G. Marques, *AI Instruments for Risk of Recidivism Prediction and the Possibility of Criminal Adjudication Deprived of Personal Moral Recognition Standards: Sparse Notes from a Layman* [w:] H. S. Antunes, P. M. Freitas, A. L. Oliveira, C. M. Pereira, E. V. de Sequeira, L. B. Xavier (red.), dz. Cyt., s. 339–351.

169 N. A. Smuha, K. Yeung, The European Union's AI Act: Beyond Motherhood and Apple Pie? [w:] N. A. Smuha (red.), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*, Cambridge University Press, 2025, s. 228 i n.

170 Zob. M. N. Dufforc, D.S. Giovanniello, *The Autonomous AI Physician: Medical Ethics and Legal Liability* [w:] H. S. Antunes, P. M. Freitas, A. L. Oliveira, C. M. Pereira, E. V. de Sequeira, L. B. Xavier (red.), dz. Cyt., s. 207–228.

6. Raportowanie i nadzór nad rozwiązaniami opartymi o AI

6.1. Metody monitorowania i oceny działania systemów AI

Wdrażanie systemów sztucznej inteligencji (AI) w praktyce niesie za sobą konieczność zapewnienia ich stałego nadzoru — nie tylko w momencie projektowania, ale również przez cały okres ich eksploatacji. Monitoring systemów AI stanowi podstawowy filar odpowiedzialnego i bezpiecznego zarządzania tymi technologiami. Jego celem jest nie tylko wykrywanie błędów technicznych czy operacyjnych, ale także ograniczanie ryzyka naruszeń praw podstawowych, przeciwdziałanie skutkom niezamierzonych decyzji oraz budowanie zaufania społecznego[171].

W szczególności monitoring umożliwia bieżącą ocenę działania algorytmów w rzeczywistym środowisku, detekcję odstępstw od oczekiwanych wzorców zachowań oraz skuteczne reagowanie na incydenty. Składa się on z dwóch uzupełniających się komponentów: zautomatyzowanego monitoringu operacyjnego, który pozwala na techniczne śledzenie stanu systemu, oraz nadzoru ludzkiego (human oversight), który gwarantuje możliwość interwencji i odpowiedzialności człowieka za decyzje podejmowane przez AI[172].

Poniżej przedstawiono szczegółowe wymagania i dobre praktyki dotyczące obu tych obszarów monitorowania, odnosząc się do uznanych standardów oraz wymogów regulacyjnych, takich jak AI Act[173].

6.1.1. Monitoring systemów AI

Monitoring systemów sztucznej inteligencji (AI) jest kluczowym elementem zapewniania ich bezpieczeństwa, niezawodności i zgodności z obowiązującymi regulacjami. Skuteczny monitoring obejmuje zarówno nadzór techniczny nad działaniem systemu, jak i udział człowieka w procesie kontroli. Wyróżniamy dwa zasadnicze komponenty monitorowania: monitoring operacyjny oraz nadzór ludzki[174].

a) Monitoring operacyjny

Monitoring operacyjny odnosi się do zautomatyzowanego nadzoru nad systemem AI w trakcie jego działania w środowisku produkcyjnym. Obejmuje on następujące elementy[175]:

- MON-01: Performance development & operation

  System AI powinien być monitorowany pod kątem wydajności i poprawności działania w rzeczywistym środowisku operacyjnym. Monitorowanie to powinno być prowadzone w sposób ciągły, umożliwiając szybkie wykrywanie anomalii.

- MON-02: Self-error detection

  System powinien posiadać zdolność do samodzielnego wykrywania błędów lub odstępstw od oczekiwanych zachowań. Funkcjonalność ta znacząco skraca czas reakcji na potencjalne problemy i zwiększa odporność operacyjną.

- MON-03: Logging

  Rejestrowanie kluczowych informacji, takich jak dane wejściowe i wyjściowe, stany modelu oraz czasy przetwarzania, jest niezbędne dla zapewnienia przejrzystości działania systemu. Narzędzia takie jak Loggly czy Tripwire

171 Grupa ekspertów wysokiego szczebla ds. SI, dz. cyt., s. 22.
172 Tamże, s. 22–23.
173 Federal Office for Information Security, *Test Criteria Catalogue for AI Systems in Finance*, Bonn 2025, s. 164 i n.
174 W „Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji" opracowanych przez Niezależną Grupę Ekspertów Wysokiego Szczebla ds. SI powołaną przez Komisję Europejską, w szczególności należy zwrócić rozdział II i III, gdzie omówiono wymogi i metody zapewniania nadzoru oraz trwałości i przejrzystości działania systemów AI. Grupa ekspertów wysokiego szczebla ds. SI, dz. cyt., s. 17 i n.
175 Federal Office for Information Security, dz. cyt., s. 164 i n.

wspierają analizę integralności logów i umożliwiają audyt.

- MON-04: Incident response procedures

  Organizacja powinna wdrożyć formalne procedury reagowania na incydenty, w tym harmonogramy interwencji oraz metody ich oceny. Skuteczność tych procedur należy regularnie testować przy użyciu kluczowych wskaźników efektywności (KPI), takich jak F1-score czy średni błąd kwadratowy (MSE).

b) Nadzór ludzki (human oversight)

Pomimo zaawansowania technologicznego, nadzór sprawowany przez człowieka pozostaje nieodzownym komponentem bezpiecznego wykorzystania systemów AI. Powinien on obejmować[176]:

- Możliwość ręcznej interwencji w działanie systemu, m.in. poprzez:
  - przyciski awaryjnego zatrzymania (emergency stop),
  - ścieżki eskalacji w przypadkach podejrzenia błędu lub ryzyka,
  - funkcje zatrzymania działania modelu i przejęcia decyzji przez człowieka (override mechanisms).

Nadzór ludzki stanowi istotny element zapewniania zgodności z zasadą proporcjonalności oraz obowiązkami wynikającymi z przepisów, takich jak AI Act. Jego obecność zmniejsza ryzyko niezamierzonych skutków decyzji podejmowanych automatycznie i zwiększa zaufanie do systemów AI w środowiskach o podwyższonym ryzyku[177].

6.1.2. Ocena działania systemów AI

Ocena działania systemów sztucznej inteligencji (AI) jest procesem wieloaspektowym, obejmującym zarówno testowanie techniczne, jak i metody o charakterze organizacyjnym i etycznym. Ma na celu zapewnienie, że system AI działa zgodnie z zamierzonymi celami, nie narusza przepisów prawa ani praw podstawowych oraz spełnia wymogi zaufania społecznego[178].

a) Testowanie i walidacja

Testowanie systemów AI powinno być prowadzone na wszystkich etapach ich cyklu życia — zarówno przed wdrożeniem (ex ante), jak i po uruchomieniu w środowisku produkcyjnym (ex post). Wskazane jest stosowanie danych realistycznych, odzwierciedlających warunki, w jakich system będzie funkcjonował w praktyce[179].

Szczególnie rekomendowane metody testowania to[180]:

- Testy kontradyktoryjne (red teaming) – polegające na aktywnym poszukiwaniu słabych punktów systemu przez ekspertów działających z perspektywy „atakującego”,
- Systemy nagród (bug bounty) – wspierające zgłaszanie wykrytych błędów przez zewnętrznych użytkowników i specjalistów, zwiększające wykrywalność nieprawidłowości i podatności.

Testowanie i walidacja powinny być zorganizowane w sposób ciągły, a ich wyniki dokumentowane i wykorzystywane w procesie doskonalenia systemu[181].

176 Federal Office for Information Security, dz. cyt., s. 130 i n.
177 Dokument „Guidelines on the definition of an AI system established by AI Act” – jest stanowiskiem podkreślającym znaczenie nadzoru i interwencji człowieka w kontekście poziomów autonomii systemów. Zob. Komisja Europejska, Annex to the Communication to the Commission Approval of the content of the draft Communication from the Commission – Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), Brussels, 6.2.2025, C (2025) 924 final.
178 W szczególności rozdział III dokumentu „Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji” zawiera opis listy kontrolnej oceny AI oraz podkreśla potrzebę wieloaspektowego podejścia do oceny systemów w całym ich cyklu życia. Grupa ekspertów wysokiego szczebla ds. SI, dz. cyt., s. 30 i n.
179 Tamże, s. 27.
180 Federal Office for Information Security, dz. cyt., s. 182.
181 Grupa ekspertów wysokiego szczebla ds. SI, dz. cyt., s. 31.

b) Lista kontrolna oceny

Jednym z praktycznych narzędzi oceny systemów AI jest lista kontrolna godnej zaufania sztucznej inteligencji. Narzędzie to pozwala usystematyzować ocenę zgodności systemu z kluczowymi wymaganiami etycznymi i operacyjnymi, takimi jak[182]:

- poszanowanie wartości i praw podstawowych,
- zapewnienie nadzoru człowieka,
- przejrzystość i możliwość wyjaśnienia działania systemu,
- odporność techniczna i bezpieczeństwo.

Lista kontrolna powinna być wdrażana w ramach procesów zarządczych, z udziałem zarówno zespołów technicznych, jak i kierownictwa organizacji. Jej stosowanie nie powinno mieć charakteru jednorazowego, lecz wspierać ciągłe doskonalenie i adaptację systemu[183].

c) Metody pozatechniczne

Ocenę systemów AI uzupełniają narzędzia pozatechniczne, których celem jest zapewnienie zgodności z obowiązującymi normami prawnymi i wartościami etycznymi. Należą do nich m.in.[184]:

- kodeksy postępowania – wewnętrzne i branżowe zasady etyczne dotyczące rozwoju i stosowania AI,
- audyty wewnętrzne i zewnętrzne – oceniające zgodność działania systemu z wymaganiami prawa i standardami organizacyjnymi,
- regulacje i normy (np. ISO/IEC, wytyczne KE) – zapewniające strukturalne ramy zarządzania ryzykiem, dokumentowania decyzji oraz rozliczalności działań.

Metody te powinny być zintegrowane z systemem zarządzania organizacją i cyklicznie aktualizowane wraz ze zmianami technologicznymi, prawnymi i społecznymi[185].

6.1.3. Narzędzia wspomagające monitorowanie i ocenę

W procesie monitorowania i oceny działania systemów sztucznej inteligencji istotną rolę odgrywają specjalistyczne narzędzia techniczne, które wspierają zarówno testowanie odporności i bezpieczeństwa, jak i ocenę skuteczności nadzoru ludzkiego. Dobór odpowiednich narzędzi powinien być dostosowany do charakteru systemu AI oraz ryzyk związanych z jego funkcjonowaniem. Poniżej przedstawiono wybrane przykłady[186]:

- **Simul8** – Narzędzie umożliwiające przeprowadzanie symulacji procesów operacyjnych z udziałem systemu AI, w tym scenariuszy związanych z nadzorem człowieka. Pozwala na ocenę skuteczności i terminowości interwencji ludzkich w przypadkach potencjalnych zagrożeń lub błędnych decyzji podejmowanych przez system.
- **OWASP ZAP**, **Burp Suite**, **Metasploit** – Zestaw narzędzi służących do testowania bezpieczeństwa interfejsów i komponentów systemu AI. Pozwalają na wykrywanie podatności w warstwie aplikacyjnej, symulowanie ataków oraz ocenę skuteczności mechanizmów kontroli dostępu i ochrony danych.
- **Gremlin** – Narzędzie wykorzystywane do testowania odporności systemów AI na nieprzewidziane zakłócenia, awarie lub błędy — tzw. chaos testing. Umożliwia celowe wprowadzanie usterek w środowisku testowym, co pozwala ocenić stabilność i odporność systemu oraz przygotowanie organizacji do sytuacji kryzysowych.

Zastosowanie tych narzędzi zwiększa wiarygodność i transparentność działania systemów AI, a także wspiera zgodność z wymaganiami normatywnymi i regulacyjnymi w zakresie zarządzania ryzykiem i bezpieczeństwem[187].

182 Tamże, s. 30.
183 Tamże.
184 Tamże, s. 31.
185 Tamże.
186 W sekcji Monitoring (rozdz. 13), Performance (rozdz. 14), oraz Human Oversight (rozdz. 11), dokumentu „Test Criteria Catalogue for AI Systems in Finance" opisano zastosowanie narzędzi takich jak Gremlin, OWASP ZAP, Burp Suite czy symulacje do oceny nadzoru. Federal Office for Information Security, dz. cyt., s. 130–189 oraz 164 i n.
187 Federal Office for Information Security, dz. cyt., s. 9–11, 139–171, 173–186.

6.1.4. Post-market monitoring (zgodność z AI Act)

Zgodnie z artykułem 72 AI Act, dostawcy systemów sztucznej inteligencji, zwłaszcza tych zakwalifikowanych jako systemy wysokiego ryzyka, są zobowiązani do prowadzenia ciągłego monitorowania ich działania po wprowadzeniu na rynek lub do użytku.

Post-market monitoring stanowi integralną część systemu zarządzania ryzykiem i ma na celu zapewnienie, że system AI nadal spełnia wymagania określone w AI Act w rzeczywistych warunkach użytkowania. Obejmuje on w szczególności[188]:

- Identyfikację nieautoryzowanych działań agentów – np. sytuacji, w których system AI działa poza zakresem przewidzianym przez dostawcę lub wykonuje działania nieprzewidziane w dokumentacji technicznej;
- Monitorowanie czasu działania bez interwencji człowieka – kluczowe dla oceny stopnia autonomii systemu oraz potrzeby zastosowania dodatkowych środków nadzoru lub ograniczenia autonomicznego funkcjonowania;
- Analizę interakcji z innymi agentami lub użytkownikami – w celu wykrycia ryzyk wynikających z nieprzewidzianej współpracy lub konfliktów pomiędzy systemami AI albo między systemem a człowiekiem.

Post-market monitoring powinien być realizowany w sposób ciągły, z użyciem odpowiednich narzędzi analitycznych, mechanizmów rejestrowania i raportowania oraz procedur szybkiego reagowania na incydenty. Jest to nie tylko wymóg prawny, ale również niezbędny element budowania zaufania do systemów sztucznej inteligencji w środowiskach rzeczywistych[189].

6.2. Rola organów nadzorczych i instytucji finansowych

6.2.1. Organy nadzorcze

Organy nadzorcze odgrywają kluczową rolę w zapewnianiu odpowiedzialnego rozwoju i stosowania systemów sztucznej inteligencji. Działają jako regulatorzy oraz strażnicy zaufania społecznego, zapewniając zgodność systemów AI z wymogami prawnymi, etycznymi i technicznymi, a także monitorując ich wpływ na prawa podstawowe i bezpieczeństwo publiczne. Ich zadania są wieloaspektowe i obejmują zarówno działania prewencyjne, jak i następcze[190].

Do głównych zadań organów nadzorczych należą[191]:

- Certyfikacja systemów SI – Organy nadzorcze weryfikują zgodność systemów AI z przyjętymi normami i standardami, w tym z normami technicznymi (np. ISO/IEC 27001, ISO/IEC 42001), wytycznymi etycznymi (np. Wytyczne KE dot. godnej zaufania AI) oraz wymogami prawnymi (w tym AI Act). Proces ten może obejmować ocenę zgodności, zatwierdzanie dokumentacji technicznej oraz ocenę systemów zarządzania jakością.
- Udział w przeglądach i środkach zaradczych. W przypadku wykrycia nieprawidłowości lub zagrożeń dla praw podstawowych, zdrowia lub bezpieczeństwa, organy nadzorcze są uprawnione do uruchamiania mechanizmów kontrolnych, przeprowadzania inspekcji oraz rekomendowania środków naprawczych, takich jak modyfikacja modelu, zawieszenie jego działania lub cofnięcie certyfikatu zgodności.
- Zapewnienie nadzoru ex ante i ex post – Organy pełnią funkcję kontrolną zarówno przed wprowadzeniem systemu AI na rynek (ex ante), jak i po jego wdrożeniu (ex post). Mogą żądać przeprowadzenia audytów wewnętrznych lub zewnętrznych, a także wymagać raportowania incydentów i wyników post-market monitoring. Szczególny nacisk kładzie się na nadzór nad systemami zakwalifikowanymi jako wysokiego ryzyka.

188 A. Oueslati, R. States-Polet, *Ahead od the Curve: Governing AI Agents Under the EU AI Act*, The Future Society 2025, s. 24.
189 Tamże.
190 Grupa ekspertów wysokiego szczebla ds. SI, dz. cyt., s. 24–26.
191 Źródłem tej treści jest dokument „Wytyczne w zakresie etyki dotyczących godnej zaufania sztucznej inteligencji", ale uzupełnieniem są zapisy AI Act, szczególnie art. 59–61 i 73–76, dotyczące uprawnień i obowiązków organów nadzorczych oraz ich roli w systemie oceny zgodności i nadzoru post-market. Tamże, s. 25–26.

- Tworzenie horyzontalnych ram oceny – Organy nadzorcze opracowują ogólne ramy oceny jakości i ryzyka systemów SI, które mogą być dostosowywane do specyfiki poszczególnych sektorów (np. transportu, ochrony zdrowia, finansów czy sektora obronnego). Celem tych działań jest zapewnienie spójności regulacyjnej, a jednocześnie zachowanie elastyczności niezbędnej do reagowania na zmiany technologiczne.

W praktyce skuteczność działania organów nadzorczych wymaga ścisłej współpracy z podmiotami notyfikowanymi, instytucjami branżowymi, twórcami technologii oraz użytkownikami końcowymi. Ich rola będzie rosła wraz z postępującym wdrażaniem AI Act i rozwojem systemów autonomicznych w kluczowych obszarach życia społeczno-gospodarczego[192].

6.2.2. Polski projekt organu nadzorczego

Na podstawie numeru wykazu UC71 w polskim Rządowym Procesie Legislacyjnym (Projekt ustawy o systemach sztucznej inteligencji), zaproponowano powierzyć określone funkcje i kompetencje Komisji Rozwoju i Bezpieczeństwa Sztucznej Inteligencji (KRiBSI). W projekcie z dnia 15 października 2024 r. zostały one precyzyjnie określone w art. 11. W związku z czym, KRiBSI pełni rolę głównego organu nadzoru rynku w rozumieniu AI Act, z szerokimi kompetencjami regulacyjnymi, nadzorczymi, edukacyjnymi i doradczymi, takimi jak:

1. Kontrola zgodności z AI Act i ustawą krajową.
2. Wspieranie rozwoju, innowacyjności i konkurencyjności w obszarze SI.
3. Reagowanie na zagrożenia bezpieczeństwa SI, w tym obsługa poważnych incydentów.
4. Opiniowanie projektów aktów prawnych dotyczących SI.
5. Wydawanie decyzji administracyjnych w sprawach naruszeń przepisów.
6. Prowadzenie działań edukacyjnych i informacyjnych.
7. Tworzenie i wspieranie piaskownic regulacyjnych.
8. Gromadzenie orzecznictwa, prowadzenie rejestru skarg i wydawanie interpretacji przepisów.

Kluczowa rola KRiBSI polega na łączeniu nadzoru prawnego, ochrony interesu publicznego i wspierania innowacji w jednym organie. W modelu proponowanym przez projekt ustawy, KRiBSI ma nie być jedynie „inspektorem" zgodności, ale strategicznym centrum kompetencji państwa w zakresie AI, podobnie jak AI Office na poziomie unijnym[193].

W kontekście przygotowanego raportu, autorzy dostrzegają w szerokim modelu funkcjonowania KRiBSI zalety, zgodne z przytoczonymi wskazaniami. Do najważniejszych należą:

- Zgodność z AI Act, tj. AI Act wymaga powołania krajowego organu nadzoru rynku zgodnie z art. 59 i 63. KRiBSI, jako jednolity punkt kontaktowy, realizuje tę funkcję zgodnie z wytycznymi Komisji Europejskiej.
- Widać w podejściu do nadzoru i innowacji ideę holistyczności. KRiBSI ma kompetencje nie tylko nadzorcze, ale również wspierające rozwój AI, co wpisuje się w koncepcję „trustworthy AI" – czyli równoważenie nadzoru, etyki i innowacyjności.
- Autorzy dostrzegają też spójność z podejściem do „governance" z poziomu UE. Takie zintegrowane działanie (compliance, rozwój, piaskownice, edukacja) jest rekomendowane m.in. przez raport „Ahead of the Curve", który podkreśla potrzebę rozdzielenia obowiązków wzdłuż łańcucha wartości, ale również wskazuje na potrzebę koordynacji i spójności regulacyjnej[194].
- Projekt jest też właściwą odpowiedzią na potrzebę koordynacji systemowej. KRiBSI jako organ nadrzędny może zapobiec fragmentacji kompetencji między UODO, UOKiK, KNF i innymi, jeśli relacje między organami zostaną wyraźnie zinstytucjonalizowane (co wybrzmiewa w wymogach dotyczących tzw. „single point of contact" w AI Act i zaleceniach ELI)[195].

192 Tamże, s. 26.
193 CEPS, …, s. 7–8, 56–57.
194 A. Oueslati, R. States-Polet, …, s. 18–26.
195 European Law Institute, …, s. 19–22 oraz 25.

W ocenie autorów aktualnie prezentowany projekt działania organu generuje też ryzyka:

- Łączenie nadzoru rynku, wydawania decyzji administracyjnych, opiniowania aktów prawnych i wspierania innowacji może prowadzić do konfliktu ról (np. „promotor rozwoju" vs „strażnik zgodności"). Autorzy postrzegają ten aspekt jako ryzyko w analizach etycznych „governance AI", bazując na idei B. C. Stahla „ekosystemowego podejścia" do etyki AI, które bazuje na elementach:
  - zróżnicowanych mechanizmów polityki publicznej,
  - zaangażowania wielu aktorów (organy nadzoru, regulatorzy, społeczeństwo),
  - przejrzystości, odpowiedzialności i partycypacji w kształtowaniu AI,
  - zachowania zgodności z koncepcją human flourishing (ludzkiego dobrostanu),
  - i unikania nadmiernej centralizacji czy technokratycznego redukcjonizmu[196].
- Efektywność takich organów jak KRiBSI zależy od zasobów: technicznych, prawnych, etycznych i organizacyjnych. W ocenie autorów utrzymanie takiego zespołu w ramach jednej instytucji jest wyzwaniem.
- Choć przy KRiBSI ma działać Rada Społeczna ds. SI, jej realny wpływ musi być zagwarantowany, aby spełnić wymóg uczestnictwa społecznego zgodnie z AI Governance typowym dla podejścia OECD i SHERPA[197].
- Bez wyraźnego przypisania ról istnieje ryzyko duplikacji albo luki nadzorczej, zwłaszcza w kontekście systemów wysokiego ryzyka (np. w finansach, ochronie zdrowia itd.).

Rola KRiBSI jako centralnego, kompetencyjnie zintegrowanego organu nadzoru rynku i innowacji jest zasadna i zgodna z AI Act oraz współczesnymi koncepcjami zrównoważonego zarządzania AI. Jednakże sama struktura nie wystarczy. Wedle autorów skuteczność KRiBSI będzie zależeć od:

- jasnego przypisania ról międzyresortowych i międzyorganowych,
- realnego zaangażowania interesariuszy (społecznych i branżowych),
- odpowiednich zasobów ludzkich i technicznych,
- przejrzystych reguł działania (np. w zakresie niezależności decyzji i funkcji kontrolnych).

6.2.3. Instytucje finansowe

Instytucje finansowe odgrywają istotną rolę w ekosystemie zarządzania ryzykiem i zgodnością systemów sztucznej inteligencji, szczególnie w kontekście stosowania AI w obszarach o wysokim wpływie na sytuację ekonomiczną klientów i stabilność rynków. Ze względu na specyfikę sektora bankowego, ubezpieczeniowego i inwestycyjnego, wdrożenie i stosowanie systemów AI musi odbywać się zgodnie z restrykcyjnymi zasadami nadzoru, audytowalności i odpowiedzialności[198]. Zgodnie z ramami ENISA dotyczącymi dobrych praktyk w zakresie cyberbezpieczeństwa dla AI, systemy AI należy traktować jako aktywa ICT, wymagające wdrożenia warstwowego podejścia obejmującego: ogólne fundamenty cyberbezpieczeństwa, specyficzne środki ochrony dla AI oraz środki sektorowe uwzględniające specyfikę rynku finansowego[199].

Ich obowiązki obejmują w szczególności[200]:

- Implementację list kontrolnych i struktur zarządzania – Instytucje finansowe powinny wdrażać listy kontrolne, procedury zgodności oraz polityki zarządzania ryzykiem dopasowane do charakteru systemów AI, z których korzystają. Dotyczy to zarówno algorytmów podejmujących decyzje kredytowe, jak i modeli wykrywających nadużycia czy analizujących zdolność płatniczą. Praktyki te powinny być zgodne z wytycznymi Komisji Europejskiej, AI Act oraz uznanymi normami (np. ISO/IEC 42001). W tym kontekście, ENISA rekomenduje trójwarstwowe podejście do zabezpieczenia systemów AI jako elementów infrastruktury ICT, obejmujące zarówno podstawowe środki cyberbezpieczeństwa, środki dedykowane AI, jak i środki sektorowe dla rynku finansowe-

196 B. C. Stahl, Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies, Leicester 2020, s.91–93, 100–102, 110 – 112.
197 Cytowana książka bazuje na wynikach projektu SHERPA (Horizon 2020), którego celem było promowanie etycznego, partycypacyjnego podejścia do AI governance — z silnym naciskiem na udział obywateli, organizacji społecznych i niezależnych ekspertów. Tamże, s. 92, 96–97, 109.
198 Federal Office for Information Security, dz. cyt., s. 9–11.
199 ENISA, Multilayer Framework ..., dz. cyt., s. 3–6.
200 Federal Office for Information Security, dz. cyt., s. 4–7 oraz 139–188.

go[201]. Instytucje finansowe mogą również czerpać z doświadczeń standardów branżowych takich jak ISO/IEC 42001 (systemy zarządzania AI) czy ISO/IEC 23894 (zarządzanie ryzykiem AI). Ich implementacja nie tylko wzmacnia zgodność z AI Act, lecz także stanowi podstawę do budowania zaufania regulatorów i klientów[202].

- Wewnętrzne procedury audytu, nadzoru i reakcji na incydenty – Instytucje mają obowiązek opracowania i stosowania mechanizmów wewnętrznej kontroli obejmujących audyty cykliczne, analizę zgodności modelu z oczekiwaniami, ocenę ryzyka systemowego oraz formalne procedury zaradcze. Wsparciem dla tych działań mogą być katalogi testowe, takie jak przedstawiony w niemieckim dokumencie „Test Criteria Catalogue for AI Systems in Finance". W przypadku bankowości i ubezpieczeń, szczególnego znaczenia nabierają mechanizmy ciągłego zarządzania ryzykiem w całym cyklu życia systemu AI, uwzględniające m.in. dynamiczne oceny zagrożeń technicznych (np. ataki typu model inversion, data poisoning) oraz zagrożeń społecznych (np. uprzedzenia algorytmiczne, brak sprawiedliwości rozkładu ryzyka kredytowego). Modele scoringowe czy systemy monitorowania nadużyć powinny być również zgodne z zasadą odporności i dokładności określoną w art. 15 AI Act oraz wspierane przez systematyczne testowanie odporności (np. AI red teaming, scenariusze kontrfaktyczne)[203].
- Zgłaszanie wyników monitorowania i zapewnienie przejrzystości – Kluczowym wymogiem jest systematyczne raportowanie wyników monitorowania działania systemów AI do właściwych organów nadzorczych oraz zapewnienie przejrzystości i identyfikowalności podejmowanych decyzji. Obejmuje to również prowadzenie dzienników decyzyjnych (decision logs), wprowadzenie mechanizmów wyjaśnialności (XAI) oraz systemu obsługi skarg.

Skuteczna realizacja tych obowiązków wymaga ścisłej współpracy między instytucjami finansowymi a organami nadzorczymi. Współpraca ta powinna odbywać się zarówno na poziomie operacyjnym (np. w ramach audytów, konsultacji i notyfikacji zmian), jak i strategicznym — poprzez wspólne tworzenie dobrych praktyk, analiz ryzyka sektorowego i rekomendacji w zakresie dalszego rozwoju i doskonalenia systemów AI[204].

6.3. Najlepsze praktyki w raportowaniu stosowania AI w płatnościach

W miarę jak systemy sztucznej inteligencji zyskują na znaczeniu w sektorze płatności, konieczne staje się wdrażanie metod nadzoru i oceny ich zgodności, które są proporcjonalne do skali i charakteru ryzyka generowanego przez te technologie. Ramy regulacyjne, takie jak AI Act, wskazują na potrzebę przyjęcia podejścia opartego na ocenie ryzyka jako podstawy do zarządzania cyklem życia systemu AI (zob. rys. 11) – od projektowania po wdrożenie i monitoring. Takie podejście pozwala organizacjom efektywnie identyfikować i klasyfikować ryzyka, dobierać adekwatne środki kontrolne, a także dokumentować ich skuteczność. W niniejszym podrozdziale omówiono kluczowe założenia tego podejścia oraz narzędzia wspierające jego realizację, w szczególności katalog kryteriów audytu AICRIV, jako przykład z Niemiec.

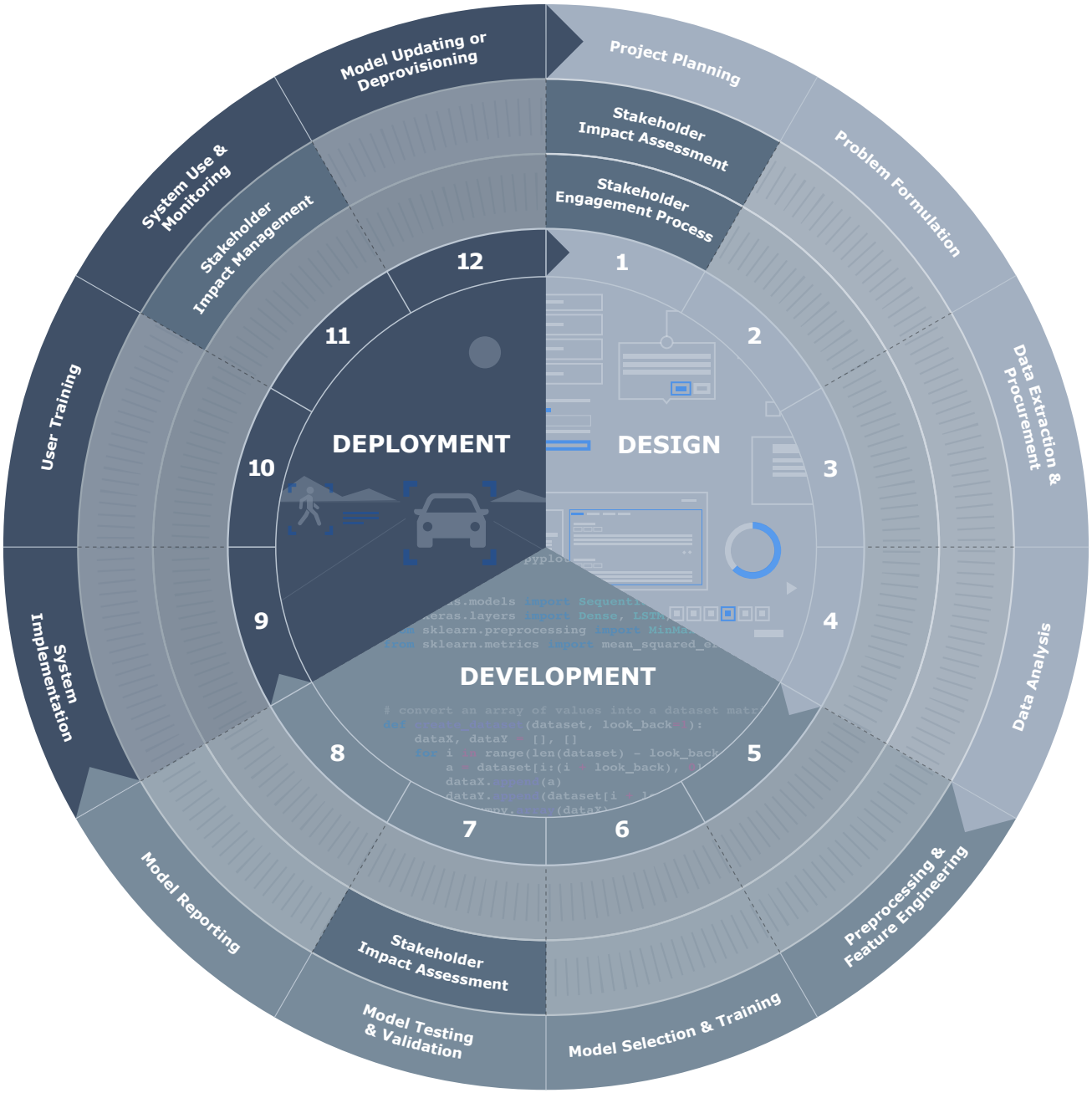
201 ENISA, Multilayer Framework ..., dz. cyt., s. 3–6.

202 Szczególnie ISO/IEC 42001 jako systemowy fundament zgodności z AI Act (zarządzanie ryzykiem, dokumentacja, audytowalność) oraz ISO/IEC 23894 jako dedykowany standard zarządzania ryzykiem dla systemów AI, relacja między wdrażaniem tych norm a oczekiwaniami regulatorów i budowaniem zaufania klientów. Obie normy wskazane są w źródłach literatury, jako kluczowe „AI-related standards" w zakresie bezpieczeństwa i zarządzania. P. Hense, T. Mustać, dz. Cyt., s. 73; ENISA, Multilayer Framework ..., dz. cyt., s. 42.

203 ENISA, Multilayer Framework ..., dz. cyt., s. 3–4 oraz 6–7, P. Hense, T. Mustać, dz. Cyt., s. 107–108, 138–139 oraz **94–98;** A. Szczęsna, M. Stachoń (red.), *Cyberbezpieczeństwo AI. AI w Cyberbezpieczeństwie*, wyd. NASK – Państwowy Instytut Badawczy, Warszawa 2023, s. 78 oraz 115.

204 Federal Office for Information Security, dz. cyt., s. 7–8 oraz 188 i n.; Grupa ekspertów wysokiego szczebla ds. SI, dz. cyt., s. 26.

Rys. 11. Cykl życia AI.



Źródło: *D. Leslie,, C. Rincón, M. Briggs, A. Perini, S. Jayadeva, A. Borda, C. SJ. Burr Bennett, M. Aitken, S. Mahomed, J. Wong, M. Waller, C. Fischer, dz. cyt, s. 44.*

6.3.1. Przyjęcie podejścia opartego na ocenie ryzyka (risk-based approach)

Wdrażanie i stosowanie systemów sztucznej inteligencji w obszarze płatności wymaga zrównoważonego podejścia do zarządzania ryzykiem, zgodnego z zasadami określonymi w unijnym rozporządzeniu AI Act oraz standardami krajowymi i branżowymi. Kluczowym narzędziem wspierającym to podejście jest katalog kryteriów audytu AI, przykładowo taki jak opracowany w ramach projektu AICRIV[205].

Podejście oparte na ocenie ryzyka zakłada dostosowanie wymagań i środków nadzorczych do konkretnego kontekstu zastosowania systemu AI. Ocenie podlega m.in. to, czy system przetwarza dane osobowe, czy stosuje modele generatywne, czy też działa w środowisku o krytycznym znaczeniu dla ciągłości działania instytucji finansowej[206].

205 Federal Office for Information Security, dz. cyt., s. 2.
206 Tamże, s. 9.

Zgodnie z katalogiem AICRIV, ryzyko systemu AI w płatnościach powinno być analizowane wieloaspektowo – technicznie (np. podatność na ataki), organizacyjnie (np. kompetencje zespołu) i etycznie (np. przejrzystość decyzji wobec klienta). Na tej podstawie dobierane są odpowiednie kryteria testowe, np. dotyczące odporności modelu, jakości danych czy transparentności działania systemu[207].

Takie podejście pozwala na prowadzenie oceny zgodności w sposób proporcjonalny, efektywny i powtarzalny, a także na identyfikację oraz zarządzanie tzw. ryzykiem resztkowym (residual risk). Ma ono również na celu wspieranie innowacyjności przy jednoczesnym zachowaniu wysokiego poziomu zaufania społecznego do systemów AI w finansach[208].

6.3.2. Raportowanie zgodne z AI Act (np. art. 17 o systemie zarządzania jakością)

Zgodnie z AI Act, organizacje wykorzystujące systemy sztucznej inteligencji, w tym w sektorze płatności, zobowiązane są do wdrożenia i utrzymywania systemu zarządzania jakością (Quality Management System – QMS). System ten powinien obejmować cały cykl życia systemu AI – od projektowania, przez rozwój, wdrożenie, aż po jego eksploatację i wycofanie z użycia.

Jednym z kluczowych elementów raportowania zgodnego z AI Act jest dokumentowanie spełnienia wymogów prawnych, w szczególności w zakresie:

- prowadzenia oceny zgodności systemu AI z wymaganiami prawnymi i technicznymi (art. 43–47 AI Act),
- systematycznego zarządzania incydentami, w tym obowiązku zgłaszania tzw. poważnych incydentów do właściwego organu nadzoru (art. 62 AI Act),
- prowadzenia rejestru dokumentacji technicznej oraz aktualizacji danych dotyczących systemu (art. 11 i 17 AI Act).

Dobrą praktyką w raportowaniu jest wdrożenie formalnej struktury QMS, która zapewnia m.in. przypisanie odpowiedzialności, wdrożenie procedur kontrolnych, weryfikację zgodności z wytycznymi UE oraz wewnętrzne audyty i przeglądy skuteczności działania systemu. W sektorze finansowym wdrożenie QMS wzmacnia zarówno zgodność regulacyjną, jak i poziom zaufania klientów oraz interesariuszy wobec wykorzystywanych rozwiązań opartych na AI.

Zgodne z art. 17 AI Act prowadzenie systemu zarządzania jakością nie tylko pozwala lepiej kontrolować ryzyka związane z eksploatacją systemów wysokiego ryzyka, lecz także wspiera przejrzystość i rozliczalność działań operatorów i dostawców AI.

6.3.3. Transparentność i wyjaśnialność decyzji AI (explainability)

Systemy sztucznej inteligencji stosowane w obszarze płatności muszą spełniać wymóg przejrzystości i wyjaśnialności (ang. explainability), który stanowi fundament godnej zaufania sztucznej inteligencji – zarówno w ujęciu etycznym, jak i regulacyjnym. AI Act, jak również wytyczne etyczne Komisji Europejskiej, podkreślają, że osoby dotknięte działaniem AI powinny mieć możliwość zrozumienia, kiedy i dlaczego system podjął określoną decyzję lub wygenerował daną rekomendację[209].

W praktyce oznacza to konieczność wdrożenia mechanizmów umożliwiających[210]:

- identyfikację sytuacji, w których decyzję podejmuje system AI, a nie człowiek,
- dostęp do uzasadnienia decyzji systemu (np. odmowy transakcji, zmiany limitu, klasyfikacji ryzyka),
- dostosowanie poziomu szczegółowości wyjaśnień do wiedzy i kompetencji odbiorcy (np. klient detaliczny vs. analityk ds. zgodności),

207 Tamże, s. 13–15, 52–54, 188–190.
208 Tamże, s. 50, 106.
209 Zob. m.in. art. 13 i 52 *AI Act*; Grupa ekspertów wysokiego szczebla ds. SI, dz. cyt., s. 16–18.
210 AI Act – w szczególności: art. 13 ust. 1 i 2, który nakłada obowiązek zapewnienia przejrzystości działania systemów wysokiego ryzyka oraz przekazywania użytkownikowi „zrozumiałych informacji" dotyczących działania systemu i sposobu jego używania, załącznik VII AI Act dotyczący systemów zarządzania jakością, który wskazuje, że systemy muszą uwzględniać m.in. sposób informowania użytkownika o ograniczeniach i przeznaczeniu AI. AICRIV w szczególności: TR-01 – ocena wymaganej wyjaśnialności; TR-02 – unikalna identyfikacja działań AI; TR-07 – dostarczanie wyjaśnień dostosowanych do poziomu odbiorcy; TR-04 – ochrona użytkownika przed nadmiernym poleganiem na AI (automation bias). Federal Office for Information Security, dz. cyt., s. 189, 191, 195 oraz 201.

- zapobieganie nadmiernemu poleganiu użytkownika na wynikach systemu (tzw. automation bias).

Zalecanym narzędziem jest wykorzystanie modeli i interfejsów wspierających tzw. wyjaśnialną sztuczną inteligencję (XAI – Explainable AI), która umożliwia zrozumiałe prezentowanie czynników wpływających na wynik modelu, bez ujawniania wrażliwych lub chronionych danych. Przykłady obejmują m.in. lokalne interpretacje (np. LIME, SHAP), graficzne reprezentacje decyzji, a także zastosowanie ustalonych reguł decyzyjnych w procesach oceniania transakcji lub ryzyka[211].

Zgodnie z katalogiem AICRIV, obowiązkiem organizacji jest również udokumentowanie, że sposób komunikowania decyzji AI jest odpowiedni z punktu widzenia przejrzystości, uczciwości oraz ochrony konsumenta – niezależnie od kanału kontaktu (np. aplikacja mobilna, system obsługi klienta, raport okresowy)[212].

6.3.4. Dokumentowanie i publikacja informacji o algorytmach

Jednym z kluczowych elementów budowania zaufania do systemów sztucznej inteligencji w sektorze płatności jest przejrzystość dotycząca ich działania. Wzorując się na brytyjskim standardzie Algorithmic Transparency Recording Standard (ATRS), zaleca się wdrażanie wewnętrznych i zewnętrznych mechanizmów dokumentowania oraz udostępniania informacji o wykorzystywanych algorytmach[213].

Standard ATRS promuje podejście polegające na rejestrowaniu[214]:

- celu i przeznaczenia danego systemu AI,
- rodzaju danych wykorzystywanych do uczenia lub działania systemu,
- podstawy prawnej i etycznej wdrożenia algorytmu,
- wpływu systemu na użytkowników – w tym potencjalnych skutków automatyzacji decyzji lub dyskryminacji,
- środków zaradczych i kontroli, które mają na celu zapewnienie nadzoru i uczciwości.

W sektorze płatności praktyka ta może obejmować publikację uproszczonych kart informacyjnych (tzw. model cards), deklaracji algorytmicznych lub wpisów do rejestrów przejrzystości prowadzonych przez regulatorów bądź instytucje branżowe. Transparentność algorytmiczna powinna również objąć zautomatyzowane decyzje mające wpływ na dostęp użytkownika do usługi płatniczej, ustalenie limitów, czy wykrywanie nadużyć[215].

Takie działania nie tylko wzmacniają zgodność z AI Act (szczególnie art. 13 i 52), ale także przyczyniają się do budowania zaufania klientów, umożliwiając im lepsze zrozumienie i egzekwowanie swoich praw.

6.3.5. Unikanie i testowanie na obecność uprzedzeń (bias)

Systemy sztucznej inteligencji stosowane w sektorze płatności muszą być projektowane i monitorowane w sposób, który minimalizuje ryzyko uprzedzeń (bias) – zarówno na poziomie danych wejściowych, jak i generowanych decyzji. Uprzedzenia te mogą prowadzić do nieuzasadnionego różnicowania traktowania użytkowników, w szczególności w zakresie oceny wiarygodności płatniczej, ryzyka nadużyć lub dostępności usług[216].

211 Tamże, dz. cyt., s. 188–201; D. Leslie,, C. Rincón, M. Briggs, A. Perini, S. Jayadeva, A. Borda, C. SJ. Burr Bennett, M. Aitken, S. Mahomed, J. Wong, M. Waller, C. Fischer, dz. cyt, s. 10–24; Grupa ekspertów wysokiego szczebla ds. SI, dz. cyt., s. 17–18.

212 Federal Office for Information Security, dz. cyt., s. 197, 207 oraz 201.

213 https://www.gov.uk/government/publications/guidance-for-organisations-using-the-algorithmic-transparency-recording-standard/algorithmic-transparency-recording-standard-guidance-for-public-sector-bodies; Zob, D. Leslie, C. Rincón, M. Briggs, A. Perini, S. Jayadeva, A. Borda, C. SJ. Burr Bennett, M. Aitken, S. Mahomed, J. Wong, M. Waller, C. Fischer, dz. cyt, s. 11–12.

214 Tamże, s. 12–13.

215 Tamże, s. 13.

216 AI Act – art. 10 ust. 2 lit. e) i art. 15 nakazują projektowanie i rozwijanie systemów wysokiego ryzyka w sposób ograniczający ryzyko błędów, uprzedzeń i dyskryminacji. Dotyczy to w szczególności systemów wykorzystywanych w ocenie ryzyka kredytowego, wykrywaniu nadużyć i innych zastosowaniach w sektorze finansowym. Ponadto Mo-

Dobrą praktyką jest przeprowadzanie systematycznych testów wykrywających uprzedzenia, z wykorzystaniem[217]:

- analizy reprezentatywności danych treningowych i operacyjnych względem populacji docelowej,
- weryfikacji zgodności z przepisami o równości traktowania i niedyskryminacji,
- symulacji wpływu polityk decyzyjnych na różne grupy odbiorców – z podziałem m.in. na płeć, wiek, status ekonomiczny, miejsce zamieszkania,
- narzędzi statystycznych i wizualizacji pozwalających zidentyfikować efekty niezamierzonych dysproporcji,
- dokumentowania wyników i zastosowanych metod korekty (np. rebalansowanie danych, algorytmiczne techniki kompensacyjne).

W katalogu AICRIV aspekt ten znajduje odzwierciedlenie w całym module FAIR – Fairness, a szczególnie[218]:

- FAIR-01 – Potential bias and impact assessment,
- FAIR-02 – Effective bias assessment,
- FAIR-03 – Bias mitigation.

Wdrożenie opisanych mechanizmów jest nie tylko elementem zgodności z AI Act (art. 10 i 15), ale także praktyczną realizacją zasady równości i uczciwości systemów decyzyjnych opartych na AI.

6.3.6. Raport końcowy z ewaluacji

Efektywne raportowanie stosowania systemów sztucznej inteligencji w sektorze płatności wymaga nie tylko przeprowadzenia testów, ale również starannego udokumentowania ich wyników. Dokumentacja ta powinna obejmować[219]:

- opis przeprowadzonych testów funkcjonalnych i niefunkcjonalnych (np. testy odporności, testy zgodności z zasadami uczciwości, testy na obecność uprzedzeń),
- wskazanie zastosowanych metryk oceny efektywności i bezpieczeństwa (np. accuracy, recall, AUC, fairness score, confidence threshold),
- wykaz użytych narzędzi i środowisk testowych (zarówno technicznych, jak i proceduralnych),
- uzasadnioną interpretację wyników, z uwzględnieniem kontekstu działania systemu,
- opis ewentualnych działań naprawczych lub kalibracyjnych wdrożonych na podstawie testów.

Katalog AICRIV jednoznacznie wskazuje, że wyniki testów oraz towarzysząca im dokumentacja powinny umożliwiać przejrzystą kontrolę zgodności i ponowną weryfikację przez niezależne podmioty – zarówno wewnętrzne jednostki audytu, jak i organy nadzoru sektorowego[220].

Takie podejście wzmacnia zaufanie do systemów AI oraz zapewnia ich rozliczalność w kontekście obowiązków wynikających z AI Act, w tym z art. 17 (system zarządzania jakością) i art. 62 (raportowanie incydentów i niezgodności).

duł FAIR (Fairness). Federal Office for Information Security, dz. cyt., s. 95–99.

217 Tamże.

218 Tamże.

219 Tamże, s. 182, 183 oraz 209.

220 Tamże, s. 106, 182 oraz 209.

7. Rekomendacje dotyczące nadzoru i raportowania systemów AI w sektorze finansowym

1. Wdrożenie kompleksowego systemu zarządzania jakością (QMS) dla AI

Instytucje korzystające z systemów AI powinny obowiązkowo wdrożyć system zarządzania jakością obejmujący cały cykl życia modelu – od projektowania po wycofanie z eksploatacji. QMS powinien zawierać procedury audytowe, identyfikację odpowiedzialności, regularne przeglądy i aktualizację dokumentacji technicznej zgodnie z art. 17 AI Act.

2. Stosowanie podejścia opartego na ryzyku (risk-based approach)

Organizacje muszą klasyfikować systemy AI pod względem ryzyka, biorąc pod uwagę charakter danych, cel zastosowania, możliwe skutki decyzji oraz ich wpływ na prawa podstawowe. Wysokie ryzyko wymaga intensywniejszego nadzoru.

3. Wdrożenie narzędzi explainable AI (XAI)

Zaleca się implementację mechanizmów wyjaśnialności takich jak LIME, SHAP czy wizualizacje reguł decyzyjnych, by użytkownicy i audytorzy mogli zrozumieć logikę działania modelu. Wyjaśnialność powinna być dostosowana do odbiorcy (klient, analityk, regulator) i uwzględniona w systemie raportowania.

4. Formalizacja nadzoru ludzkiego (human oversight)

Każdy system AI powinien mieć jasno określone ścieżki interwencji manualnej. Wymaga to wdrożenia funkcji „emergency stop", mechanizmów „override" oraz protokołów eskalacji. Personel powinien być przeszkolony w zakresie odpowiedzialności i reagowania na nieprawidłowości.

5. Ciągły monitoring post-market

Organizacje muszą prowadzić ciągły monitoring działania systemów po ich wdrożeniu, z wykorzystaniem narzędzi analitycznych i logów operacyjnych. Monitoring powinien obejmować analizę czasu autonomicznego działania systemu, nieautoryzowanych działań oraz interakcji z innymi agentami.

6. Stosowanie testów kontradyktoryjnych i scenariuszy awaryjnych

Rekomenduje się cykliczne przeprowadzanie testów red teaming, symulacji chaos testing oraz scenariuszy kontrfaktycznych, które pozwalają ocenić odporność systemu na ataki, błędy danych i nieprzewidziane sytuacje.

7. Audyt i dokumentacja uprzedzeń algorytmicznych

Systemy AI muszą być testowane pod kątem obecności uprzedzeń (bias), w tym uprzedzeń danych, strukturalnych i decyzyjnych. Należy prowadzić dokumentację analizy reprezentatywności danych, skutków decyzji oraz stosować środki korekty, zgodnie z modułem FAIR w AICRIV.

8. Zastosowanie branżowych standardów zgodności

Wdrożenie norm ISO/IEC 42001 (zarządzanie AI) i ISO/IEC 23894 (zarządzanie ryzykiem AI) powinno stać się standardem dla instytucji finansowych. Ich przestrzeganie wzmacnia pozycję organizacji wobec organów nadzoru oraz stanowi dowód zgodności z AI Act.

9. Tworzenie i publikacja kart modelowych (model cards)

Zaleca się opracowywanie kart informacyjnych dla każdego systemu AI zawierających jego cel, typ modelu, źródła danych, ryzyka i zastosowane środki kontrolne. Karty mogą być publikowane w rejestrach transparentności

10. Integracja AI z istniejącymi systemami compliance

Systemy AI powinny być zintegrowane z politykami zgodności i procedurami compliance obowiązującymi w instytucji. Dotyczy to m.in. przeciwdziałania praniu pieniędzy, przeciwdziałania oszustwom czy ochrony danych osobowych (RODO). AI nie może funkcjonować jako „czarna skrzynka".

11. Raportowanie incydentów i prowadzenie dzienników decyzyjnych

Organizacje muszą wdrożyć obowiązek zgłaszania poważnych incydentów zgodnie z art. 62 AI Act oraz prowadzić rejestry decyzji podejmowanych przez AI (decision logs). Informacje te powinny być dostępne dla audytorów oraz organów nadzoru.

12. Szkolenia i kompetencje zespołów wdrażających AI

Wdrożenie AI wymaga interdyscyplinarnego zespołu z wiedzą z zakresu analizy danych, regulacji prawnych, etyki technologii i bezpieczeństwa IT. Instytucje powinny inwestować w szkolenia z zakresu zgodności z AI Act, monitorowania ryzyk oraz etycznego projektowania algorytmów.


PROGRAM
ANALITYCZNO
BADAWCZY