

Bezpieczeństwo i analiza ryzyka we wdrażaniu sztucznej inteligencji w bankowości detalicznej

Dariusz Fatuła

dr hab., prof. UAFM, Wydział Zarządzania, Mediów i Technologii, Uniwersytet Andrzeja Frycza Modrzewskiego w Krakowie (Faculty of Management, Media and Technology, Andrzej Frycz Modrzewski Krakow University, Kraków, Poland)
<https://orcid.org/0000-0002-6704-2585>

Wprowadzenie

Sztuczna inteligencja (AI) od wielu lat stanowi jeden z najbardziej dynamicznie rozwijających się obszarów technologii, które wywierają istotny wpływ na sektor bankowości detalicznej. Pierwsze zastosowania AI w bankach, sięgające lat 80. i 90. XX wieku, obejmowały systemy eksperckie i proste modele predykcyjne wspierające analizę ryzyka, ocenę zdolności kredytowej oraz wykrywanie prób oszustw (Baesens i in., 2016). Dzięki postępowi w dziedzinie uczenia maszynowego, przetwarzania dużych zbiorów danych oraz rosnącej dostępności mocy obliczeniowej, systemy te zostały znacznie rozbudowane i obecnie stanowią podstawę wielu procesów decyzyjnych w bankowości detalicznej. Sztuczna inteligencja w bankowości detalicznej przestaje być eksperymentem technologicznym, a staje się infrastrukturą decyzyjną o znaczeniu systemowym. W polskiej i międzynarodowej literaturze opisuje się już szerokie spektrum zastosowań – od automatyzacji procesów operacyjnych, przez ocenę zdolności kredytowej, wykrywanie nadużyć i personalizację ofert, po wsparcie nadzoru zgodności regulacyjnej. Równolegle podkreśla się, że każdy z tych obszarów generuje nowe typy ryzyka oraz wzmacnia istniejące ekspozycje sektora finansowego

(Szostek i in. 2022; Ziółkowska 2023; Shi i in. 2022; Noriega i in. 2023). W analizach dotyczących wpływu sztucznej inteligencji na rynek finansowy akcentuje się dualny charakter tej technologii: z jednej strony poprawia ona efektywność, dokładność predykcji i dostępność usług, z drugiej – zwiększa złożoność systemu, tworzy zależność od dostawców zewnętrznych (w tym Big Tech), pogłębia koncentrację rynku i rodzi trudne do oszacowania ryzyka dla stabilności finansowej (Financial Stability Board, 2024). W polskich opracowaniach kierowanych do sektora bankowego zwraca się uwagę, że instytucje finansowe stają wobec równoległego wymogu budowania przewagi konkurencyjnej dzięki AI oraz wdrożenia rygorystycznych mechanizmów kontroli, aby uniknąć naruszeń praw konsumentów, kar finansowych i utraty reputacji (Centrum Procesów Bankowych i Informacji, 2024).

Celem artykułu jest przegląd literatury oraz raportów z badań w celu pogrupowania i usystematyzowania wspomnianych zagadnień ryzyka wdrażania AI w bankowości detalicznej w obszarach łączących perspektywę zarządzania, prawa i technologii. Podjęto też próbę identyfikacji najważniejszych luk pomiędzy teorią a praktyką oraz zaproponowano ramy koncepcyjne ich wypełnienia.

Najważniejsze zastosowania AI w bankowości detalicznej

AI znajduje coraz szersze zastosowanie w bankowości detalicznej, transformując praktycznie wszystkie aspekty usług bankowych oferowanych klientom indywidualnym. Jednym z najstarszych i najbardziej zaawansowanych zastosowań AI w bankowości detalicznej jest automatyczna ocena kredytowa (*credit scoring*) i oceny ryzyka klienta (*underwriting*), gdzie systemy AI wykorzystują uczenie maszynowe do analizy historii kredytowej, zachowań finansowych, danych transakcyjnych oraz alternatywnych źródeł danych (takich jak np. historia płatności za media lub zakupy w sieci) w celu ustalenia zdolności kredytowej klienta (Sachan i in. 2020; Nimma-gadda, 2022).

Chatboty oparte na AI oraz wirtualni asystenci stanowią transformacyjną technologię w obsłudze klientów banków detalicznych. Banki wdrażają zaawansowane systemy konwersacyjne, które mogą obsługiwać miliony zapytań klientów. Wirtualni asystenci wykorzystują przetwarzanie języka naturalnego (*natural language processing*, NLP) do wyjaśniania warunków usług finansowych, pomagania w inicjowaniu transakcji, oferowania spersonalizowanych produktów oraz wspierania w zarządzaniu kontem.

AI znacznie podnosi jakość personalizacji usług bankowych poprzez analizę historii transakcji, zachowań wydatkowych oraz preferencji klientów. Systemy rekomendacyjne mogą automatycznie identyfikować momenty, w których klient jest gotów do przejścia na wyższą kategorię karty kredytowej, zaciągnięcia pożyczki czy otwarcia rachunku oszczędnościowego. AI służy też prognozowaniu, którzy klienci

mogą opuścić bank (*customer churn*), na podstawie analizy wzorców aktywności, zmian w wydatkach oraz interakcji z bankiem. Pozwala to bankom na proaktywne podejmowanie działań zapobiegawczych (retencyjnych). Modele predykcyjne segmentują również klientów, przewidują odejścia, analizują lojalność oraz generują wartościowe spostrzeżenia (*insighty*) na potrzeby marketingu i zarządzania portfelem produktów.

Systemy AI do wykrywania oszustów znacznie ograniczają koszty związane wyłudzeniami, oszustwami i innymi bezprawnymi działaniami w bankowości (Uddin 2025). Algorytmy uczenia maszynowego analizują w czasie rzeczywistym miliony transakcji, identyfikując anomalie i podejrzanе wzorce, które mogą wskazywać na oszustwa lub nieautoryzowany dostęp. Banki wdrażają zaawansowane systemy AI, które zmniejszają liczbę fałszywych alarmów (*false positives*), jednocześnie poprawiając detekcję rzeczywistych zagrożeń. Aplikacje wykorzystują zaawansowane modele statystyczne i generatywne do aktywnej identyfikacji podejrzaných płatności oraz wczesnego ostrzegania o zagrożeniach.

Przestrzeganie zgodności z szybko zmieniającymi się przepisami prawa (*compliance*) w bankowości ewoluuje dzięki sztucznej inteligencji. Skomplikowane regulacje jak AI Act czy Basel IV wymagają aktualizacji bądź zmiany wewnętrznych przepisów i procedur bankowych. AI szybciej niż prawnicy wyłapuje niezgodności w wielu skomplikowanych dokumentach, opisach, materiałach marketingowych i proponuje odpowiednie zmiany.

AI przetwarza ogromne zbiory danych w czasie rzeczywistym, identyfikując ryzyka operacyjne, kredytowe, rynkowe i regulacyjne szybciej niż tradycyjne metody, automatyzuje złożone procesy, zwiększa precyzję predykcji i zapewnia zgodność z obowiązującym prawem. Biorąc pod uwagę powyższe zastosowania, można uznać, że AI jest obecnie ważnym narzędziem zarządzania szeroko pojętym ryzykiem w działalności banku. Równocześnie jednak samo wdrażanie i wykorzystywanie AI rodzi pewne specyficzne ryzyka.

Obszary ryzyka we wdrażaniu i funkcjonowaniu AI

Ryzyka prawne i regulacyjne

Wdrożenie systemów AI w bankowości detalicznej generuje ryzyka prawne i regulacyjne wynikające z nieprzejrzystości algorytmów oraz trudności w zapewnieniu kontroli człowieka nad decyzjami automatycznymi. EU AI Act klasyfikuje systemy oceny zdolności kredytowej jako „wysokiego ryzyka”, nakładając obowiązek wdrożenia systemów zarządzania ryzykiem, przejrzystości decyzji oraz dokumentacji cyklu życia modelu, z karami do 35 mln euro lub 7% globalnego obrotu rocznego (Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689, 2024).

General Data Protection Regulation (GDPR, w Polsce: Rozporządzenie o Ochronie Danych Osobowych – RODO), obowiązujące od 25 maja 2018 r. w całej UE, w art. 22 wymaga prawa do wyjaśnienia decyzji automatycznych, jednak mechanizmy takie jak SHAP¹ czy LIME² nie spełniają prawnych wymogów „wyjaśnienia”, generując jedynie techniczne przybliżenia nieprzystające do oczekiwań regulacyjnych (Edwards i Veale, 2017). Kroll i in. (2017) wskazują, że pełna transparentność kodu źródłowego algorytmów prowadzi do paradoksu: ujawnienie umożliwia obchodzenie systemu przez zainteresowanych lub osłabia konkurencyjność twórców. Proponują zamiast tego kryptograficzne mechanizmy procedury zgodności wyników z polityką decyzyjną bez ujawniania algorytmu. Citron i Pasquale (2014) dokumentują nieprzejrzystość systemów oceny kredytowej (*scoring*), gdzie instytucje finansowe odmawiają dostępu do szczegółów metodologii, uniemożliwiając audyt indywidualnych decyzji, co prowadzi do arbitralnych ocen kredytowych. Szostek i in. (2022) analizują specyficzne ryzyka dla polskiego sektora bankowego, wskazując na odpowiedzialność banków za systemy AI stron trzecich, trudności w kwalifikacji modeli jako wysokiego ryzyka oraz wymogi dokumentacji zgodne z AI Act. Financial Stability Board (2024) podkreśla, że rosnąca zależność od zewnętrznych dostawców AI zwiększa koncentrację ekspozycji, gdzie awarie lub błędy u jednego dostawcy mogą wywołać efekty systemowe. Podczas gdy istnieje znacząca literatura dotycząca akceptacji technologii oraz zaufania do automatyzacji (Lee i See, 2004; Venkatesh i in., 2003), brakuje szczegółowych badań empirycznych na temat długoterminowego wpływu interakcji z systemami AI na zachowania finansowe klientów banków detalicznych. Raport World Economic Forum (2025) wskazuje, że wymaga się większej wiedzy na temat stosunku pracowników do AI oraz dynamiki zmian miejsc pracy w kontakcie z AI. Brakuje również badań na temat psychologicznych skutków stosowania AI przez pracowników i klientów. Pytania: jak klienci reagują na wielokrotne błędy AI? Czy generatywna AI faktycznie buduje zaufanie, czy je podważa? — pozostają w dużej mierze niezbadane. Brakuje też empirycznych badań na temat tego, w jaki sposób łączne zastosowanie Blockchain i AI mogłoby zwiększyć bezpieczeństwo transakcji i zapobiegać oszustwom w bankach detalicznych, szczególnie w kontekście zapewniania audytowalności i transparentności rozwiązań informatycznych (Pramudito i in., 2025).

Ryzyko operacyjne i automatyzacji procesów

Ryzyko operacyjne definiowane jest jako potencjalne straty wynikające z nieadekwatnych lub nieskutecznych procesów wewnętrznych, ludzi oraz systemów lub

¹ SHAP (SHapley Additive exPlanations) działa na zasadzie wartości Shapleya z teorii gier – krok po kroku wyjaśnia, ile każda cecha przyczyniła się do konkretnej predykcji modelu.

² LIME (Local Interpretable Model-agnostic Explanations) działa lokalnie – krok po kroku aproksymuje skomplikowany model (*black-box*) prostym modelem liniowym w otoczeniu konkretnej predykcji.

z wydarzeń zewnętrznych (McLemore i Mihov 2026). W bankowości detalicznej automatyzacja procesów za pomocą sztucznej inteligencji – obejmująca m.in. poznanie swojego klienta (KYC – *know your customer*), przeciwdziałanie praniu pieniędzy (AML – *anti-money laundering*), detekcję nadużyć (*fraud detection*), ocenę kredytową (*scoring*) i obsługę klienta przez rozmowy z AI – zmienia profil tego ryzyka poprzez zwiększoną złożoność systemów, zależność od dostawców zewnętrznych oraz nowe możliwości ataków hakerskich przez sieć kontaktu klienta z instytucją finansową. Jednym z najważniejszych ryzyk w tym obszarze jest tzw. stronniczość lub w kontekście społecznym uprzedzenie (*bias*) modeli wynikające z doboru danych wejściowych, które z różnych przyczyn zawierają wyniki preferujące lub częściej odrzucające wybranych klientów ze względu na ich cechy jak wiek, płeć, miejsce zamieszkania (Akinrele, 2025). Generatywna AI (GenAI) może udzielać klientom błędnych porad finansowych lub wprowadzać w błąd co do oferty, co prowadzi do strat finansowych i utraty zaufania (halucynacje i błędy modeli). Degradacja jakości modelu w czasie (np. gdy zmieniają się zachowania klientów), prowadzi do błędnych decyzji kredytowych lub związanych z wykryciem oszustwa (fraudowych), jeśli model nie jest stale rekalkulowany. Poleganie na kilku dużych dostawcach chmury/AI stwarza zagrożenie systemowe w przypadku awarii ich infrastruktury (koncentracja ryzyka kontrahenta). Ryzyko kontrahenta polega przede wszystkim na potencjalnych stratach wynikających z niewywiązania się z zobowiązań przez zewnętrznych dostawców rozwiązań AI, fintechy czy platformy technologiczne. Opóźnienia w dostawie, uaktualnieniach czy poprawie błędów systemu, nie mówiąc już o bankructwie lub zaprzestaniu działalności przez kontrahenta, może narazić bank na dezorganizację pracy, utratę zaufania i wysokie kary nadzoru. Nawet w pozornie dobrze działających systemach dostarczonych przez kontrahentów mogą występować błędy i uprzedzenia decyzyjne oraz możliwość wycieku danych i utraty prywatności, co stanowi osobne ryzyko opisane dalej.

Na poziomie strategicznym minimalizacja tych ryzyk wymaga określenia tolerancji wobec błędów oraz uprzedzeń, zatwierdzenia budżetu na audyty i szkolenia pracowników oraz ustanowienie struktury, np. komitetu ds. etyki AI, który nadzorować będzie zgodność z wytycznymi wewnętrznymi. Na poziomie operacyjnym powinno się opracować procedury testowania uprzedzeń w modelach oceny kredytowej – np. przez badanie, czy prawdopodobieństwo odrzucenia kredytu nie różni się istotnie między różnymi grupami. Zespoły odpowiedzialne za analizę danych (*data science*) i uczenie maszynowe (*machine learning*) powinny zadbać o reprezentatywność danych treningowych (uwzględnienie różnych grup demograficznych), zastosować techniki zapobiegania stronniczości (np. *reweighting*, *adversarial debiasing*) w procesie modelowania, przeprowadzić testy diagnostyczne uprzedzeń (*bias testing*) przed i po wdrożeniu, oraz dokumentować ograniczenia modelu w odniesieniu do różnych segmentów klientów.

Ryzyko wycieku danych i utraty prywatności

AI wymaga dostępu do dużych wolumenów wrażliwych danych, których niewłaściwa ochrona prowadzi do wycieków i naruszeń prywatności (Häuselmann, 2024). Wyzwania te są potęgowane przez wymogi GDPR/RODO i inne reżimy prawne narzucające konieczność wyjaśniania procesu decyzji przez AI (Wachter i in., 2017). Ponadto generatywne modele AI mogą zapisywać (zapamiętywać, memoryzować) wrażliwe dane z danych treningowych i reprodukować je niezamierzenie (Fredrikson i in., 2015). Zgodnie z prawem do wyjaśnienia (*right to explanation*), wprowadzonym przez RODO (art. 22), banki muszą mieć zdolność wyjaśnienia każdej decyzji algorytmicznej wpływającej na klienta. To wymaga nie tylko przejrzystych modeli, ale także stworzenia procedur i nadzorowania procesów zbierania, przechowywania i przetwarzania danych oraz otrzymanych wyników, które mogą wytrzymać audyt regulacyjny.

Ryzyka etyczne i społeczne

Ryzyka te mają swoje źródła w innych, wyżej już opisanych ryzykach, ale ich znaczenie jest nieco inne wymiarze społecznym i etycznym (Mäntymäki 2022). W erze mediów społecznościowych, informacja o powtarzającym się dyskryminowaniu określonych grup klientów banku rozprzestrzenia się wirusowo. Halucynowanie AI w zakresie doradztwa dla klienta może narazić bank na śmieszność i zarzuty braku profesjonalizmu. Wyciek danych klientów, oprócz aspektów prawnych, może spowodować utratę wiarygodności i zaufania, co jest niezwykle groźne w sektorze bankowym. Informacje o podatności na ataki hackerskie, o przerwach w elektronicznym kontakcie z klientami, o zawieszaniu się stron internetowych, szybko roznoszą się po sieci społecznościowej i budzą czasem przesadne obawy wśród klientów. Utrata zaufania może spowodować odpływ klientów, prowadzący w skrajnych wypadkach do utraty płynności. Konieczne w takich wypadkach będą skoordynowane działania działów marketingowych, prawnych i technicznych banku.

Ryzyko kadrowe i luka kompetencyjna

Brak pracowników potrafiących nadzorować i weryfikować działanie zaawansowanych modeli AI (*human-in-the-loop*) znacząco zwiększa ryzyko błędów operacyjnych. Wprowadzanie AI do działalności przedsiębiorstw i banków stało się krytycznym wyzwaniem dla kapitału ludzkiego (Nagarakanti, 2025). Narracja medialna często koncentruje się na utracie miejsc pracy. Nie jest to jednak jedyny efekt – ważniejszy leży w niedopasowaniu kompetencyjnym oraz zjawisku utraty tradycyjnych umiejętności pracowników pozostających w organizacji (Tshewang, 2024). Pracownicy mogą zacząć bezkrytycznie ufać wynikom generowanym przez AI (np. ocenie zdolności kredytowej lub rekomendacji inwestycyjnej), nie rozumiejąc probabilistycznej natury modelu. Pracownicy nieświadomi zjawiska halucynacji (w GenAI) lub uprzedzeń

(w modelach ML – *machine learning*) mogą nieumyślnie wprowadzać klientów w błąd lub naruszać regulacje, używając narzędzi AI w sposób niezgodny z przeznaczeniem. Personel polegający na automatyzacji traci zdolność do samodzielnego wykonywania złożonych zadań. Młodszy personel, który od początku kariery korzysta z automatycznych systemów (problem pokolenia Z), może nigdy nie nabyć głębokiej intuicji kredytowej ani zrozumienia fundamentów ryzyka, które posiadali ich poprzednicy. W sytuacji awarii systemu AI lub wystąpienia nietypowego przypadku, tacy pracownicy stają się bezradni operacyjnie. W przypadku konieczności powrotu do procedur manualnych (np. w wyniku cyberataku na infrastrukturę AI), zarządzający bankiem mogą odkryć, że nie posiadają już kadr zdolnych do ręcznego procesowania transakcji czy oceniania ryzyka w akceptowalnym czasie.

Współpraca na linii człowiek–AI generuje nowe typy błędów operacyjnych. Wprowadzenie „copilotów” (AI Copilots) do oddziałów i call center zmienia dynamikę pracy. Pracownicy mogą mieć tendencję do faworyzowania sugestii algorytmu nad własny osąd lub sprzeczne dowody, co jest szczególnie niebezpieczne w AML (*anti-money laundering*, przeciwdziałanie praniu pieniędzy) i *fraud* (wyłudzeniach danych, pieniędzy). Paradoksalnie, AI może zwiększyć obciążenie kognitywne pracowników. Zamiast wykonywać proste, powtarzalne czynności (które przejęło AI), pracownicy muszą stale zajmować się wyłącznie trudnymi, wyjątkowymi przypadkami, które odrzucił automat, co prowadzi do szybszego zmęczenia, a w konsekwencji wypalenia zawodowego. Luka kompetencyjna potęguje ryzyko operacyjne, wizerunkowe i tworzy asymetrię między szybkością wdrażania algorytmów a zdolnością personelu do ich bezpiecznej obsługi i nadzoru.

Opisane ryzyka nie wyczerpują wszystkich zagrożeń związanych z AI w tym obszarze. Identyfikacja ryzyk jako podstawa do wdrażania odpowiednich metod i technik zarządzania, powinna brać pod uwagę występujące luki badawcze. Najważniejsze z nich to: niedostatek długoterminowych studiów empirycznych nad interakcjami wielu modeli AI (tzw. bias kumulacyjny) i manipulacją algorytmami oraz brak kompleksowych badań nad erozją zaufania klientów wskutek błędów modeli i nad samym zastosowaniem AI. Świadomość takich luk powinna wyczulić menadżerów na obserwację nietypowych, nieopisanych zjawisk zachodzących w środowisku nadzorowanej pracy.

Dobrym i stosunkowo prostym narzędziem wskazującym pilność podejmowania działań przez zarządzających jest mapa ryzyk. We wdrażaniu AI w bankowości detalicznej wizualizować można zagrożenia według dwóch wybranych osi, np.: prawdopodobieństwa wystąpienia (niskie–wysokie) i wpływu na operacje lub klientów (niski–wysoki), dzieląc je na strefy: zieloną (niska pilność działania), żółtą (średnia) i czerwoną (wysoka). Przykładowo cyberataki na modele AI z wyciekiem danych mają wysokie prawdopodobieństwo wystąpienia i wysoki wpływ na opinie klientów (strefa czerwona). Brak zgodności dokumentów i procedur z regulacjami ma średnie

prawdopodobieństwo wystąpienia i stosunkowo niski wpływ na klientów. Opóźnienia we wdrażaniu nowych systemów ma wysokie prawdopodobieństwo wystąpienia, ale niski lub średni wpływ na operacje i klientów (strefa żółta). Niskie prawdopodobieństwo i niski wpływ (strefa zielona) mogą mieć np. większe błędy w raportach wewnętrznych (łatwe do wykrycia na wielu etapach) i niedostosowanie starszych pracowników do specyfiki pracy z AI (możliwość szkoleń, przesunięcie na stanowiskach).

Zarządzanie ryzykiem we wdrażaniu i funkcjonowaniu AI

O ile w latach 2020–2023 wdrażanie systemów uczących się (ML) oraz generatywnej sztucznej inteligencji (GenAI) miało charakter punktowy i eksperymentalny, o tyle w dwu kolejnych latach algorytmy przestały pełnić jedynie funkcję narzędzi wspierających analitykę marketingową czy proste interakcje z klientem (chatboty). Stały się nieodłącznym elementem procesów decyzyjnych o wysokim ryzyku, przede wszystkim takich jak ocena zdolności kredytowej (*credit scoring*), detekcja nadużyć w czasie rzeczywistym (*anti-fraud*) oraz automatyzacja procesów przeciwdziałania praniu pieniędzy (AML). Regulacje prawne na poziomie Unii Europejskiej, jak Rozporządzenie o sztucznej inteligencji (EU AI Act), a w Polsce Rekomendacje H i W Komisji Nadzoru Finansowego (KNF), powodują, że skuteczne wdrożenie AI w bankowości detalicznej nie jest już wyzwaniem tylko technologicznym, lecz przede wszystkim zarządczym i prawnym.

Jako przykłady wdrażania AI do wykrywania przestępstw finansowych można podać bank HSBC, który w 2024 r. dzięki AI analizował ok. 900 milionów transakcji miesięcznie na 40 milionach kont klientów (Calvery, 2024). Przykładem wykrycia występowania ryzyk wdrażania AI w bankowości detalicznej jest stronniczość algorytmów wykazana przez badaczy z Lehigh University. Duże modele językowe (LLM) częściej odmawiały kredytów i nakładały wyższe oprocentowanie na wnioski od czarnoskórych klientów w porównaniu do identycznych profili białych klientów, szczególnie przy wyższym ryzyku (Bowen III i in., 2024).

Podczas gdy skuteczność predykcyjna modeli AI rośnie dzięki zastosowaniu głębokich sieci neuronowych, ich transparentność i wyjaśnialność dla organów nadzorczych maleje. W bankowości detalicznej, która z definicji opiera się na zaufaniu klienta masowego oraz ścisłej ochronie konsumenta, ta asymetria informacyjna staje się ryzykiem systemowym. Tradycyjne podejście do zarządzania ryzykiem modelu (*model risk management* – MRM), znane z Rekomendacji W Komisji Nadzoru Finansowego, okazuje się niewystarczające. Klasyczne MRM koncentrowało się na stabilności statystycznej modeli ekonometrycznych. Zarządzanie ryzykiem AI wymaga natomiast podejścia socjotechnicznego, uwzględniającego nie tylko poprawność matematyczną, ale także sprawiedliwość (*fairness*), wyjaśnialność (*explainability*) oraz odporność (*robustness*) na ataki (*adversarial attacks*), mające na celu

oszukiwanie modeli AI poprzez m.in. subtelnie (ale często złośliwie) modyfikowane dane wejściowe.

Fundamentem ładu korporacyjnego w bankowości, szczególnie w aspekcie AI, jest model Trzech Linii Obrony (*Three Lines of Defense* – 3LoD), który wywodzi się z tradycyjnej doktryny audytu i zarządzania ryzykiem operacyjnym (Schuett, 2025). Choć nieformalne struktury podziału odpowiedzialności istniały w bankowości od dekad, model ten został formalnie skodyfikowany i spopularyzowany przez Instytut Audytorów Wewnętrznych (The Institute of Internal Auditors – IIA)³. Model IIA z 2013 r. (przywoływany w literaturze naukowej z zakresu finansów, zarządzania i audytu jako fundamentalny model teoretyczny stosowany w praktyce gospodarczej) *The Three Lines of Defense in Effective Risk Management and Control* (The Institute of Internal Auditors, 2013) zawierał odpowiedź na kryzys finansowy z lat 2007–2008, który obnażył strukturalne słabości w nadzorze bankowym – sytuacje, w których „wszyscy odpowiadali za ryzyko, więc nikt za to nie odpowiadał”. Model został szybko zaadaptowany przez Bazylejski Komitet Nadzoru Bankowego (Basel Committee on Banking Supervision – BCBS) jako globalny standard zarządzania ryzykiem operacyjnym. W klasycznym ujęciu model ten zakłada hierarchiczną separację funkcji:

- I linia (poziom operacyjny): zarządzający ryzykiem na co dzień w toku działalności operacyjnej dostarczają produktu/usługi, bezpośrednio zarządzając ryzykiem;
- II linia (ryzyko, zgodność, kontrola finansowa): wyznacza ramy, polityki i narzędzia do monitorowania ryzyka oraz rzuca wyzwanie I linii (testowanie pod kątem wybranych wskaźników⁴), raportuje zbiorcze ryzyko do zarządu;
- III linia (audyt wewnętrzny): zapewnia niezależną, obiektywną ocenę skuteczności działania dwóch pierwszych linii, raportując bezpośrednio do Rady Nadzorczej.

W literaturze przedmiotu (Tammenga, 2020) podkreśla się ryzyko „audytu fasadowego”. Występuje ono wtedy, gdy audytorzy sprawdzają jedynie istnienie dokumentacji („czy istnieje dokument opisujący model?”), a nie weryfikują faktycznej skuteczności modelu. W dobie AI III linia musi ewoluować w stronę tzw. audytu algorytmicznego. Polega on na tym, że audytorzy III linii muszą być w stanie niezależnie pobrać próbkę danych i przetestować model (*re-performance*). Następnie powinni zweryfikować, czy człowiek w pętli decyzyjnej (*human-in-the-loop*) faktycznie

³ W 2020 r. Instytut Audytorów Wewnętrznych opublikował zaktualizowaną wersję tego modelu pod nową nazwą, w której pominięto człon „obrony”: *The IIA's three lines model: An update of the three lines of defense*. (The Institute of Internal Auditors, 2020). Nowe podejście zamiast sztywnej obrony kładzie większy nacisk na:

- współpracę i komunikację między liniami zamiast tworzenia izolowanych „silosów”,
- wsparcie realizacji celów, a nie tylko ochronę przed stratami,
- wykorzystanie technologii, w tym AI i automatyzacji, do ciągłego monitorowania ryzyk w czasie rzeczywistym.

⁴ II linia odpowiada również za interpretację regulacji (np. EU AI Act) i przełożenie ich na konkretne metryki techniczne, które I linia musi spełnić (np. wskaźnik *fairness* na poziomie minimum 80%).

nadzoruje AI, czy tylko bezrefleksyjnie zatwierdza jego decyzje (zjawisko *automation bias*). Docelowo powinni ocenić kulturę ryzyka w organizacji: czy presja na wynik biznesowy (I linia) nie prowadzi do ignorowania ostrzeżeń generowanych przez systemy monitorujące. Inne wyzwanie, dotyczące głównie II linii, polega na luce kompetencyjnej. Tradycyjni pracownicy (głównie prawnicy) sprawdzający zgodność z przepisami (*compliance*) często nie posiadają wiedzy technicznej, by ocenić ryzyko związane z parametrami sieci neuronowej. Dlatego w nowoczesnym modelu 3LoD, II linia musi zostać wyposażona w kompetencje techniczne. Wskazane jest zatem zatrudnianie lub/i przesuwanie do zespołu „walidatorów technicznych” – najlepiej pracowników posiadających kompetencje w zakresie analityki danych (*data scientists*).

Opisany powyżej model Trzech Linii przeszedł drogę od koncepcji teoretycznej do standardu praktyki zawodowej. KNF w swoich rekomendacjach – przede wszystkim w H (system kontroli wewnętrznej), ale pośrednio i w innych: D (testy i odporność), M (automatyzacja ryzyka), W (walidacja i monitoring) – wprost wymaga od banków posiadania struktury opartej na Trzech Liniach. Komitet Bazylejski włączył ten model do globalnych wytycznych nadzorczych dla sektora bankowego, co uczyniło go standardem globalnym. W najbliższych latach najważniejszymi dokumentami w zakresie zarządzania ryzykiem i bezpieczeństwem w bankach staną się standard ISO/IEC No. 42001:2023 (International Organization for Standardization, 2023) oraz Artificial Intelligence Risk Management Framework (National Institute of Standards and Technology, 2023).

Pierwszy dokument został opublikowany 18 grudnia 2023 r. jako norma międzynarodowa i jest rodzajem warstwy dla tworzenia polityk zarządczych i wyznaczania odpowiedzialnych za ryzyka – ustanawia komitety sterujące i zarządza przeprowadzanie audytów wewnętrznych (odpowiednik linii II i III w modelu Trzech Linii). Drugi standard (styczeń 2023) jest dokumentem rządowym ze Stanów Zjednoczonych, skąd pochodzą najważniejsze i największe systemy AI, nie ma więc w Europie i w Polsce znaczenia obowiązującego, ale stanowi rodzaj podręcznika dla projektantów systemów i informatyków dostarczających rozwiązania AI w bankowości (odpowiednik linii I w modelu Trzech Linii).

Norma ISO 42001 nie skupia się wyłącznie na inżynierii modelu ML, lecz na całościowym systemie zarządzania organizacją (AIMS – Artificial Intelligence Management System). Oznacza to przesunięcie punktu ciężkości z poprawności kodu na odpowiedzialność organizacyjną, co dobrze wpisuje się w wymogi EU AI Act oraz oczekiwania organów nadzoru finansowego (w Polsce KNF). Norma składa się z 10 klauzul głównych (obligatoryjnych) oraz czterech załączników o charakterze normatywnym i informacyjnym. Przykładowo klauzula 4 wymusza m.in. identyfikację interesariuszy: klientów (obawiających się dyskryminacji), regulatora oraz dostawców technologii. Klauzula 6 wymaga przeprowadzenia oceny ryzyka AI (*AI risk assessment*) w kontekście niezamierzonego wyniku działania modelu

(np. halucynacji GenAI) lub szkody społecznej (np. wykluczenia finansowego). Klauzula 7 wymusza szkolenia nie tylko dla inżynierów, ale także dla personelu odpowiedzialnego za kontakt z klientem (np. doradców klienta korzystających z rekomendacji AI), aby uniknąć zjawiska bezkrytycznego ufania maszynie. Odpowiada to w pewnym stopniu drugiej linii obrony w modelu Trzech Linii. Klauzula 8 nakłada rygorystyczne wymogi co do kontroli dostawców zewnętrznych, za których błędy bank (instytucja kupująca usługę) odpowiada jak za błędy własne. Klauzula 9 wprowadza wymóg ciągłego monitorowania modelu AI pod kątem jego „zdrowia” (np. *data drift*, spadek dokładności) i raportowania wyniku do zarządu (II linia obrony). Audyt wewnętrzny (III linia obrony) musi okresowo weryfikować zgodność systemu z normą i raportować do organu wyższego (np. rady nadzorczej). Załączniki zawierają listę celów stosowania zabezpieczeń, określają konieczność sformalizowania zasad etycznych – nie tylko deklaratywnie, ale także w postaci procedury operacyjnej. Określają kontrolę projektowania (np. czy system ma prawo odmówić kredytu samodzielnie?), cyklu życia, weryfikacji i walidacji (tzw. V&V) – testów przed zatwierdzeniem i wdrożeniem modelu do działania (II linia). Kolejne załączniki określają kontrolę danych dla modeli AI (wymagają sprawdzenia, czy dane treningowe nie zawierają historycznych uprzedzeń), wprowadzenia procedur transparentności i wyjaśnialności decyzji dla klientów. Podsumowując, norma ta stanowi pomost między wizją technologii (inżynierów ML), a światem regulacji (prawników) i wdrażania (zarządu, nadzoru), jej wdrożenie jest dowodem dojrzałości cyfrowej organizacji. W erze mediów cyfrowych, gdy klienci boją się dehumanizacji bankowości, znak jakości ISO 42001 pełnić może także funkcję marketingową. Jest komunikatem, że algorytmy banku są bezpieczne, audytowane i pod kontrolą. W relacjach B2B certyfikat może być warunkiem koniecznym współpracy.

Drugi wspomniany dokument powstał w 2023 r. w Narodowym Instytucie Standaryzacji i Technologii (NIST) przy Departamencie Handlu USA. Formalnie jest to standard dobrowolny może on jednak pełnić rolę „miękkiego prawa” w sektorze finansowym, stanowiąc merytoryczne dopełnienie dla „twardych” regulacji takich jak EU AI Act czy norma ISO/IEC 42001. Ogólna idea tego dokumentu (podobnie jak wspomnianych wcześniej regulacji i norm) wskazuje na odejście od traktowania systemu AI jako czysto matematycznego modelu (kodu), na rzecz analizy jego interakcji z otoczeniem społecznym. W bankowości oznacza to, że błąd algorytmu nie jest mierzony tylko stratą finansową banku, ale także szkodą wyrządzoną klientowi (*harm to individuals*), grupie społecznej (*harm to groups*) lub rynkowi (*harm to society*). Struktura tego dokumentu opiera się na czterech wzajemnie przenikających się funkcjach:

1. Zarządzaj strategicznie (*govern*) – w ramach tej funkcji wyższe kierownictwo instytucji finansowej ustala apetyt na ryzyko, buduje kulturę organizacyjną promującą bezpieczeństwo oraz zapewnia zasoby niezbędne do skutecznego nadzoru.

To tutaj tworzone powinny być polityki określające, jakie systemy AI są dopuszczalne do wdrożenia, a jakie naruszają standardy etyczne banku. Kluczowym zadaniem funkcji *govern* jest również ustanowienie jasnych struktur odpowiedzialności, eliminujących zjawisko rozmycia winy w przypadku błędów algorytmicznych, oraz integracja zarządzania ryzykiem AI z istniejącymi systemami prawnymi i cyberbezpieczeństwa.

2. Mapuj (*map*) – funkcja ta polega na systematycznym wykrywaniu zależności, ograniczeń technologii oraz założeń przyjętych na etapie projektowania, które w rzeczywistym świecie mogą okazać się błędne. Kluczowym aspektem jest tu analiza całego cyklu życia systemu – od doboru danych treningowych, przez wdrożenie, aż po wycofanie z eksploatacji – w celu przewidzenia potencjalnych negatywnych skutków, zanim one wystąpią.
 3. Mierz (*measure*) – funkcja polega na zastosowaniu metod ilościowych i jakościowych do oceny zidentyfikowanych wcześniej ryzyk, z wykorzystaniem rygorystycznych testów i audytów technicznych. W ramach tej funkcji organizacja wdraża konkretne metryki do oceny takich parametrów jak sprawiedliwość algorytmiczna (*fairness*), wyjaśnialność decyzji czy odporność na ataki, i monitoruje zachowanie modelu w warunkach kontrolowanych oraz produkcyjnych. Pomiar ten nie może być tylko jednorazowym aktem, lecz ciągłym procesem, który ma na celu wykrycie degradacji jakości modelu lub pojawienia się nowych zagrożeń w czasie.
 4. Zarządzaj operacyjnie (*manage*) – funkcja ta obejmuje wdrażanie technicznych i organizacyjnych mechanizmów kontrolnych, takich jak korekta algorytmów, czyszczenie danych czy wprowadzenie nadzoru ludzkiego nad decyzjami maszyny. W sytuacjach gdy ryzyka nie da się skutecznie ograniczyć, faza ta legitymizuje decyzję o rezygnacji z wdrożenia systemu lub jego wyłączeniu. Zarządzanie operacyjne obejmuje również ciągły monitoring powdrożeniowy, reagowanie na incydenty oraz komunikację z interesariuszami na temat zidentyfikowanych zagrożeń. W odniesieniu do powyższych funkcji NIST zdefiniował siedem cech jakości systemów AI:
1. Poprawny i niezawodny (*valid & reliable*): system działa zgodnie z założeniami nie tylko w laboratorium, ale na rzeczywistych danych produkcyjnych, które są często zaszumione.
 2. Bezpieczny (*safe*): system nie stwarza zagrożenia dla życia, zdrowia ani mienia. W bankowości oznacza to np. stabilność systemu transakcyjnego sterowanego AI.
 3. Zabezpieczony i odporny (*secure & resilient*): odporność na ataki typu zatrucie danych treningowych (*data poisoning*) czy kradzież własności intelektualnej modelu (*model extraction*).
 4. Rozliczalny i przejrzysty (*accountable & transparent*): możliwość odtworzenia ścieżki decyzyjnej. Kto, kiedy i dlaczego zatwierdził daną wersję modelu?

5. Wyjaśnialny i interpretowalny (*explainable & interpretable*): możliwość zrozumienia decyzji przez człowieka. Jest to wymóg prawny (RODO) w przypadku decyzji automatycznych.
6. Chroniący prywatność (*privacy-enhanced*): wykorzystuje takie narzędzia i techniki, które trenują modele bez ujawniania wrażliwych danych klientów.
7. Sprawiedliwy (*fair with harmful bias managed*): aktywnie zarządza uprzedzeniami historycznymi i społecznymi zawartymi w danych.

Warto podkreślić, że NIST traktuje te cechy jako wzajemnie powiązane, choć czasem pozostające w konflikcie. W praktyce zarządzania bezpieczeństwem wdrażania i funkcjonowania AI w bankowości detalicznej oba opisane dokumenty pełnią już ważną rolę i będą miały coraz większe znaczenie. Dokument NIST dostarcza precyzyjnego języka, taksonomii ryzyk oraz konkretnych metryk do pomiaru sprawiedliwości czy odporności algorytmów. Jest on swoistą „instrukcją obsługi” regulacji, pozwalającą przetłumaczyć abstrakcyjne pojęcia prawne na procedury testowe. Z kolei ISO 42001 uzupełnia ten obraz o warstwę zarządczą, oferując strukturę pozwalającą na integrację nadzoru nad AI z istniejącymi procesami korporacyjnymi, co czyni bezpieczeństwo procesem audytowalnym i zrozumiałym dla zarządów.

Polska, jako członek Unii Europejskiej, podlegać będzie przepisom EU AI Act od 2 sierpnia 2026 r. Zarządzający bankami powinni więc przygotować się z wyprzedzeniem (stworzyć formalne struktury zarządzania AI i wyszkolić pracowników), zamiast czekać na sankcje regulacyjne. Inwestycja w szeroko pojęte bezpieczeństwo AI na etapie wdrażania jest znacznie tańsza niż retrospektywne naprawy po incydentach lub sankcjach nadzorczych. Kluczowe jest powołanie zespołów i określenie ich kompetencji i odpowiedzialności, np. wg opisanego powyżej modelu Trzech Linii. Upraszczając, powinny to być co najmniej trzy zespoły:

1. nadzoru nad AI, który zdefiniuje zasady bezpieczeństwa, skoordynuje procesy zarządzania i rozwoju wszystkich systemów, zapewni zgodność z celami biznesowymi i przepisami prawa, zintegruje z politykę wykorzystania AI z innymi systemami banku;
2. testerów systemów AI, odpowiadający za każdy system AI w jego całym cyklu życia – od planowania do wycofania; zespół ten powinien odpowiadać za specyfikację przypadków użycia, ocenę ryzyka, dokumentowanie parametrów, monitoring wydajności, a także kontrolę jakości każdego wdrożonego modelu;
3. inżynierów poszczególnych systemów, odpowiedzialny bezpośrednio za każdy algorytm (zarówno wewnętrznie rozwinięty, jak i pozyskany od zewnętrznych dostawców); zespół ponosi odpowiedzialność za ciągły monitoring behawioralny algorytmu, jego bieżącą konfigurację, testowanie, weryfikację.

Role te powinny być formalnie zdefiniowane w strukturze organizacyjnej oraz wpisane w systemy odpowiedzialności i motywacji pracowników.

Najmniej rozpoznane w literaturze przedmiotu i w zasadzie niezbadane na większych i różnorodnych grupach pozostaje opisane wcześniej ryzyko kadrowe i luka kompetencyjna (Kowalik i in., 2025; Bell i in., 2025). W tym obszarze banki powinny:

- prowadzić systematyczną diagnostykę i szkolenia rozwijające kompetencje pracowników w obszarach wykorzystujących AI,
- przeddefiniować role pracowników – z bezkrytycznych operatorów systemów na nadzorców i interpretatorów AI,
- utrzymywać i okresowo ćwiczyć „manualne” procedury, aby w razie awarii AI lub cyberataku organizacja nie utraciła zdolności operacyjnych,
- budować ścieżki kariery i systemu premiowego dla ról związanych z AI.

Podobnie słabo zbadany pozostaje obszar reakcji klientów banków na coraz szersze wdrażanie systemów AI. Banki w tym zakresie powinny (Bharti i in., 2023; Naveen i in., 2026):

- przeprowadzać reprezentatywne badania ankietowe na klientach ($n > 500$), mierząc postawy wobec AI w różnych obszarach ich styku z technologią,
- przeprowadzać wywiady pogłębione ($n = 20$ do 30) z klientami ze zróżnicowanych segmentów (młodzi, starsi, częściowo/całkowicie wykluczeni itd.), identyfikujące obawy wobec AI w sposób jakościowy,
- prowadzić badania w obszarze *user experience* (np. *eyetracking*, analiza map ciepła, testy A/B),
- wdrażać pilotażowe wersje usług AI (chatboty, robodoradcy) dla podgrup klientów,
- mierzyć behawioralne wskaźniki po kontakcie z AI: np. satysfakcji, lojalności, rezygnacji (*churn*),
- analizować dane behawioralne: np. czas sesji, liczba sesji porzuconych, konwersja zapytań na transakcje,
- segmentować klientów, np. wg demografii, częstotliwości korzystania, zaufania do technologii.

Wymienione powyżej badania i wnioski z nich płynące tradycyjnie pozostają w gestii szeroko pojętych działów/komórek marketingu. W obszarze AI powinny jednak być konsultowane z omówionymi wcześniej strukturami odpowiedzialnymi za bezpieczeństwo i wdrażanie systemów AI.

Podsumowanie

Coraz szersze wykorzystanie sztucznej inteligencji we wszystkich obszarach naszego życia sprawia, że w bankowości detalicznej nie wystarczy tylko adaptować nowe narzędzia, aby nadążyć za konkurencją. Konieczne jest wprowadzenie dużych zmian organizacyjnych. Działania te mogą opierać się na gotowych ramach z aktów prawnych, wytycznych branżowych i literatury naukowej – ale sukces wymagać będzie

także własnej inicjatywy w adaptacji do specyfiki lokalnego rynku. Opisane tu i w innych publikacjach naukowych i branżowych rozwiązania zapewnić mogą fundament, ale banki muszą dostosować swoje działania do własnej specyfiki: posiadanych zasobów, zbadanych luk (np. kompetencyjnych) i lokalnego rynku (klientów). Twórcze rozwiązania powinny powstać w ramach interdyscyplinarnych zespołów skupiających naukowców i informatyków analizujących duże zbiory danych (*data scientists*), prawników, psychologów behawioralnych, marketingowców i managerów specjalizujących się w strukturach bankowych. Wydaje się, że oprócz konkurencji o klienta (co można na pewnym etapie pozostawić działom marketingu) banki mogłyby też współpracować na poziomie stworzenia specyficznych standardów, np. w formie międzybankowego konsorcjum (z KNF, ZBP), np. do walidacji modeli AI na polskich danych. Taka inicjatywa może być pomostem pomiędzy regulacjami skupiającymi się na zgodności, wytycznymi branżowymi skupiającymi się na obsłudze technologii i pracami naukowymi operującymi na teorii. Banki potrzebują praktycznych narzędzi z danymi lokalnymi. Koszty badań i wdrożeń powinny zwrócić się przez uniknięcie kar i wzrost zaufania klientów (Bharti i in., 2023). Kluczowym zadaniem dla zarządzających ryzykiem we wdrażaniu AI jest integracja wszystkich wymiarów i obszarów przedstawionych w artykule oraz zastosowanie branżowych standardów zarządzania. Menadżerowie muszą bezustannie śledzić nowe technologie, pojawiające się standardy, analizować bieżące przypadki zagrożeń w swojej pracy, tak aby szybko – samodzielnie, w zespołach lub raportując do specjalistów i wyższych szczebli zarządzania – wypełniać luki badawcze między teorią a praktyką.

Bibliografia

- Akinrele, O. (2025). Bias mitigation in AI-driven banking: Legal, ethical, and regulatory perspectives. *International Journal of Science and Research Archive*, 17(2), 373–379. <https://doi.org/10.30574/ijrsra.2025.17.2.3013>
- Baesens, B., Roesch, D., Scheule, H. (2016). *Credit risk analytics: Measurement techniques, applications, and examples in SAS*. John Wiley & Sons.
- Bell, S., Gadanecz, B., Gambacorta, L., Perez-Cruz, F., & Shreeti, V. (2025, 24 kwietnia). Artificial intelligence and human capital: Challenges for central banks, *BIS Bulletin*, 100. <https://www.bis.org/publications/bulletin-100-artificial-intelligence-and-human-capital-challenges-central-banks.pdf>
- Bharti, S. S., Prasad, K., Sudha, S., & Kumari, V. (2023). Customer acceptability towards AI-enabled digital banking: A PLS-SEM approach. *Journal of Financial Services Marketing*, 28(4), 779–793. <https://doi.org/10.1057/s41264-023-00241-9>
- Bowen III, D. E., McKay Price, S., Stein, L. C. D., Yang, K. (2024, 3 czerwca). Measuring and mitigating racial disparities in LLMs: Evidence from a mortgage underwriting experiment. SSRN. <http://dx.doi.org/10.2139/ssrn.4812158>
- Calvery, J. (2024, 10 czerwca). *Harnessing the power of AI to fight financial crime*. HSBC. <https://www.hsbc.com/news-and-views/views/hsbc-views/harnessing-the-power-of-ai-to-fight-financial-crime>

- Centrum Procesów Bankowych i Informatyki. (2024). *Era AI: jak sztuczna inteligencja zmienia bankowość?* Wydawnictwo Centrum Procesów Bankowych i Informatyki. https://bank.pl/wp-content/uploads/2024/12/Raport_ERA_AI.pdf
- Citron D. K., Pasquale F. (2014). The scored society: Due process for automated predictions. *Washington Law Review*, 89(1), 1–34.
- Edwards, L., Veale, M. (2017). Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. *Duke Law & Technology Review*, 16(1), 18–84. <https://scholarship.law.duke.edu/dltr/vol16/iss1/2>
- Financial Stability Board (2024, 14 listopada). *The financial stability implications of artificial intelligence*. <https://www.fsb.org/uploads/P14112024.pdf>
- Fredrikson, M., Jha, S., Ristenpart, T. (2015, 12–16 października). Model inversion attacks that exploit confidence information and basic countermeasures. *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (pp. 1322–1333). <https://doi.org/10.1145/2810103.2813677>
- International Organization for Standardization. (2023). *Information technology — Artificial intelligence — Management system* (ISO/IEC Standard No. 42001:2023). <https://www.iso.org/standard/81230.html>
- Häuselmann, A. N. (2024, 23 kwietnia). *EU privacy and data protection law applied to AI: unveiling the legal problems for individuals* [Doctoral thesis, Leiden University]. <https://scholarlypublications.universiteitleiden.nl/handle/1887/3747996>
- Kowalik, Z., Lewandowski, P., Geodecki, T., & Grodzicki, M. (2025). Automation in shared services centres: Implications for skills and autonomy. *The Economic and Labour Relations Review*, 36(2), 563–581. <https://doi.org/10.1017/elr.2025.10026>
- Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, 165(3), 633–705.
- Lee, J. D., See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Mäntymäki, M., Minkinen, M., Birkstedt, T., Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics*, 2, 603–609. <https://doi.org/10.1007/s43681-022-00143-x>
- McLemore, P., Mihov, A. (2026). AI and operational losses: Evidence from U.S. Bank Holding Companies. *The Review of Corporate Finance Studies*, artykuł cfag003. <https://doi.org/10.1093/rcfs/cfag003>
- Nagarakanti, R. C. (2025). Human-AI collaboration in financial services: Augmenting decision-making with cloud-native intelligence. *European Journal of Computer Science and Information Technology*, 13(22), 23–41. <https://doi.org/10.37745/ejcsit.2013/vol13n222341>
- National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. <https://doi.org/10.6028/NIST.AI.100-1>
- Naveen, K. R., Janani, M., Pavithra, R. (2026). Assessing the influence of AI adoption and trust on customer satisfaction in digital banking services. *SAM Advanced Management Journal*, 91(2), 165–183. <https://doi.org/10.1108/SAMAMJ-12-2024-0100>
- Nimmagadda, V. S. P. (2022). Artificial Intelligence for automated loan underwriting in banking: Advanced models, techniques, and real-world applications. *Journal of Artificial Intelligence Research and Applications*, 2(1). <https://aimlstudies.co.uk/index.php/jaira/article/view/204>
- Noriega, J. P., Rivera, L. A., Herrera, J. A. (2023). Machine learning for credit risk prediction: A systematic literature review. *Data*, 8(11), artykuł 169. <https://doi.org/10.3390/data8110169>

- Pramudito, D., Na'am, J., Ernawan, F. (2025). Exploring blockchain and AI in digital banking: A literature review on transactions enhancement, fraud detection, and financial inclusion. *Sistemasi: Jurnal Sistem Informasi*, 14(3), 1448–1459. <https://doi.org/10.32520/stmsi.v14i3.5231>
- Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji). *Dziennik Urzędowy Unii Europejskiej*, Dz.U. L 2024/1689. <https://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX:32024R1689>
- Sachan, S., Yang, J. B., Xu, D. L., Benavides, D. E., Li, Y. (2020). An explainable AI decision-support-system to automate loan underwriting. *Expert Systems with Applications*, 144, artykuł 113100. <https://doi.org/10.1016/j.eswa.2019.113100>
- Schuett, J. (2025). Three lines of defense against risks from AI. *AI & Society*, 40, 493–507. <https://doi.org/10.1007/s00146-023-01811-0>
- Shi, S., Tse, R., Luo, W., D'Addona, S., Pau, G. (2022). Machine learning-driven credit risk: a systemic review. *Neural Computing and Applications*, 34(17), 14327–14339. <http://doi.org/10.1007/s00521-022-07472-2>
- Szostek, D., Bar, G., Prabucki, T., Nowakowski, M. (2022, sierpień). *Zastosowanie sztucznej inteligencji w bankowości – szanse oraz zagrożenia. Analiza prawnoregulacyjna wpływu technologii uczenia maszynowego i pokrewnych na obowiązki sektora bankowego z zakresu zapewnienia zgodności (compliance) oraz zarządzania ryzykiem*. Raport opracowany na zlecenie Programu Analityczno-Badawczego Fundacji Warszawski Instytut Bankowości.
- Tammenga, A. (2020). The application of Artificial Intelligence in banks in the context of the three lines of defence model. *Maandblad voor Accountancy en Bedrijfseconomie*, 94(5/6), 219–230. <https://doi.org/10.5117/mab.94.47158>
- The Institute of Internal Auditors. (2013). The Three Lines Of Defense In Effective Risk Management And Control [Position paper]. <https://theiia.fi/wp-content/uploads/2017/01/pp-the-three-lines-of-defense-in-effective-risk-management-and-control.pdf>
- The Institute of Internal Auditors. (2020). *The IIA's Three Lines Model: An update of the Three Lines of Defense* [Position paper]. <https://www.theiia.org/globalassets/documents/resources/the-iias-three-lines-model-an-update-of-the-three-lines-of-defense-july-2020/three-lines-model-updated-english.pdf>
- Tshewang S. (2024, 6 kwietnia), Impact of artificial intelligence on job market in the banking sector: A focus on the Indian context. SSRN. <http://dx.doi.org/10.2139/ssrn.4786492>
- Uddin, N. (2025). Role of AI in preventing financial crime: A comprehensive analytical review. *Journal of Economic Criminology*, 13, artykuł 100200. <http://doi.org/10.1016/j.jeconc.2025.100200>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <http://doi.org/10.2307/30036540>
- Wachter, S., Mittelstadt, B., Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, 7(2), 76–99. <http://doi.org/10.1093/idpl/ix005>
- World Economic Forum. (2025). *Future of Jobs Report 2025*. <https://www.weforum.org/reports/the-future-of-jobs-report-2025>

Ziółkowska, E. (2023), Wpływ sztucznej inteligencji na rynek finansowy w procesie podejmowania decyzji ekonomicznych – szanse, wyzwania i rekomendacje, *Studia i Prace Kolegium Zarządzania i Finansów*, 192, 89–108. <https://doi.org/10.33119/SIP.2023.192.5>

Bezpieczeństwo i analiza ryzyka we wdrażaniu sztucznej inteligencji w bankowości detalicznej

Streszczenie

W artykule omówiono najważniejsze zastosowania sztucznej inteligencji w bankowości detalicznej, wskazano ryzyka we wdrażaniu i funkcjonowaniu AI. Zidentyfikowano luki pomiędzy teorią a praktyką oraz zaproponowano ramy koncepcyjne ich wypełnienia. W zakresie zarządzania ryzykiem przedstawiono model Trzech Linii, zestawiając go z normą ISO/IEC 42001:2023 oraz standardem opublikowanym przez Narodowy Instytut Standaryzacji i Technologii (NIST) przy Departamencie Handlu USA. Następnie zaproponowano, jak wybrane koncepcje teoretyczne, normy i zalecenia przełożyć na tworzenie i zarządzanie strukturami nadzorującymi funkcjonowanie AI w banku, tak aby były zgodne z wchodzącym w pełni w życie od sierpnia 2026 r. unijnym aktem w sprawie sztucznej inteligencji (EU AI Act). Wskazano też na przyszłe działania poprawiające bezpieczeństwo systemu branży bankowej. Innowacyjne rozwiązania mogą nie tylko służyć konkurencji poszczególnych banków, ale także w pewnym zakresie wspierać współpracę w dziedzinie szeroko pojętego bezpieczeństwa.

Słowa kluczowe: sztuczna inteligencja, bankowość detaliczna, zarządzanie ryzykiem, zarządzanie bezpieczeństwem

Security and risk analysis in the implementation of artificial intelligence in retail banking

Abstract

The article discusses the most important applications of artificial intelligence in retail banking, identifies risks in the implementation and operation of AI. It highlights gaps between theory and practice and proposes conceptual frameworks to address them. In the area of risk management, it examines the Three Lines model, juxtaposing it with the standard ISO/IEC 42001:2023 and the standard published by the National Institute of Standards and Technology (NIST) at the U.S. Department of Commerce. Then it proposes how selected theoretical concepts, standards, and recommendations can be translated into creating and managing structures overseeing the operation of AI in a bank, ensuring compliance with the EU AI Act, which comes fully into force in August 2026. It also points to future actions that improve the safety of the banking sector system. Innovative solutions can not only serve the competition of individual banks but also, to some extent, support cooperation in the broadly understood area of security.

Keywords: artificial intelligence, retail banking, risk management, security management

Wkład autorów / Author Contributions

Konceptualizacja / Conceptualization: D.F.

Metodologia / Methodology: D.F.

Analiza formalna / Formal analysis: D.F.

Badania / Investigation: D.F.

Zasoby / Resources: D.F.

Przygotowanie wersji wstępnej / Writing – original draft: D.F.

Recenzja i edycja / Writing – review & editing: D.F.

Materiały dodatkowe / Supplementary Materials

Nie dotyczy / Not applicable.

Finansowanie / Funding

Badania nie otrzymały żadnego zewnętrznego finansowania. / This research received no external funding.

Oświadczenie o zatwierdzeniu etycznym / Ethics Approval Statement

Nie dotyczy / Not applicable.

Oświadczenie o świadomej zgodzie / Informed Consent Statement

Nie dotyczy / Not applicable.

Oświadczenie o dostępności danych / Data Availability Statement

Udostępnianie danych nie ma zastosowania do tego artykułu, ponieważ w ramach badania nie wygenerowano ani nie analizowano żadnych nowych danych. / Data sharing is not applicable to this article as no new data were generated or analyzed in this study.

Oświadczenie o użyciu generatywnej sztucznej inteligencji / Declaration of Generative AI Use

Do przygotowania niniejszego artykułu nie wykorzystano narzędzi generatywnej sztucznej inteligencji. / No generative artificial intelligence tools were used in the preparation of this manuscript.

Podziękowania / Acknowledgments

Nie dotyczy / Not applicable.

Konflikt interesów / Conflict of Interest

Dariusz Fatuła deklaruje brak konfliktu interesów. / Dariusz Fatuła declares no conflict of interest.

