Universität Zürich[UZH]

University of Zurich

Department of Finance

Advancing Risk Management in Swiss Banking: AI-driven Client Documentation, Identity Matching, and Fraud Detection

April 23, 2026

Editors: Prof. Dr. Walter Farkas[1,2], Dr. Daniel Fasnacht[1]
Scientific Coordinator: Patrick Lucescu[1]
Contributors Chapter 1: Venkata Akhila Rani Obilisetty[1], Oyelana Emmanuel[1], Tosino Alessandro Pietro[3], Hatem Khrouf[1]
Contributors Chapter 2: Antonio Del Rio[1], Florian Kuonen[4], Stefan Uzelac[3]
Contributors Chapter 3: Tamanna Kumavat[1]

[1]University of Zurich
[2]ETH Zürich
[3]Universität Liechtenstein
[4]University of Basel

Abstract

Artificial intelligence is rapidly transforming financial services, driven by rising regulatory expectations, digital client engagement, and increasingly sophisticated financial crime. Swiss banks face persistent challenges in documentation quality, identity verification across fragmented data sources, and the detection of evolving fraud typologies. Traditional control frameworks are often insufficient to manage these risks at scale and in real time, creating the need for more continuous, data driven, and intelligence-based approaches to risk management.

This white paper from the Department of Finance at the University of Zurich examines how AI can address these industry gaps, drawing on the outcomes of RiskON 2025, an initiative that connects academic research with real world banking challenges. Multidisciplinary teams developed solutions across three core domains: client documentation, adverse media identity matching, and AI driven fraud detection. The proposed approaches demonstrate how advanced analytics can strengthen documentation standards, improve identity assurance, enhance fraud detection, and support more scalable and transparent compliance processes.

The findings underline the strategic importance of close collaboration between academia, financial institutions, regulators, and technology partners. Structured experimentation and practical solution development within open innovation frameworks are critical to accelerating the integration of advanced analytics into operational risk and compliance processes. Looking ahead, this collaborative model can strengthen the global competitiveness of the Swiss financial center, reinforce trust, and position Switzerland as a leading hub for responsible and resilient AI enabled financial services.

Contents

Prologue

Risk management in financial services is entering a period of structural transformation. The growing complexity of regulatory frameworks, the digitalization of client interactions, and the rapid evolution of artificial intelligence are fundamentally reshaping how financial institutions detect, assess, and mitigate risk. Traditional control architectures, often built on static rules, periodic reviews, and fragmented data, are increasingly insufficient in an environment defined by real time information flows, cross border operations, and rapidly evolving technological threats. As a result, risk management is moving beyond reactive compliance toward continuous and intelligence driven oversight.

Addressing this challenge requires new forms of collaboration that extend beyond traditional institutional boundaries. Effective innovation in regulated environments depends on closer integration between academia, industry, and technology providers, bringing together theoretical rigor, operational expertise, and scalable technological capabilities.

The RiskON initiative was founded in 2023 to meet this need. It stands as a distinctive initiative bridging the worlds of academic research and financial industry practice. Under the leadership of the University of Zurich's (UZH) Department of Finance, with Professor Dr. Walter Farkas and Dr. Daniel Fasnacht at the helm, and supported by the UZH Innovation Hub in collaboration with the N9 House of Innovation, RiskON has established itself as a dynamic platform where theoretical knowledge is put to work on real-world risk management problems. By bringing together university researchers and prominent Swiss financial institutions, RiskON functions as an incubator for applied, AI-driven innovation in banking. The initiative translates emerging technologies into practical concepts, prototypes, and governance frameworks, contributing to the development of more adaptive, resilient, and future-oriented risk management practices.

This report aims to document and critically examine the challenges posed during RiskON 2025 and to explore the range of innovative solutions developed by the participating student teams.

The event follows a two days hackathon-style format, structured around an intensive and highly focused programme. Participants engage in opening sessions, team pitch presentations, challenge-focused hackathon sprints, and networking and

awards activities. Operating within the framework of the Innosuisse AI Innovation Booster, RiskON accelerates the translation of AI research into real-world financial applications through open innovation. The programme provides a collaborative environment in which multidisciplinary teams develop and refine AI-based solutions, with the ambition of producing minimum viable products (MVPs) or proofs of concept (PoCs) within six months of the event.

At the heart of the RiskON initiative lies its challenge-based competition model. For the 2025 edition, three leading Swiss financial institutions — Bank Julius Bär, Zürcher Kantonalbank (ZKB), and EFG Bank — each contributed a strategically motivated research challenge sitting at the intersection of emerging AI technologies and pressing risk management needs. The challenges brought together over 50 students from nearly all Swiss universities, as well as the University of Liechtenstein. Participants represented a wide range of disciplines, including mathematics, data science, economics, quantitative finance, law, and were enrolled in master, PhD, and executive education programmes. These diverse and multidisciplinary teams engaged with the following complex, industry-defined problems:

- *Quality Assurance for Client Contact Notes — How can AI Help?* (Bank Julius Bär),
- *KYC vs. Adverse Media — Can AI Match Identities Across Sources?* (Zürcher Kantonalbank),
- *AI-powered Fraud Detection — How Can Patterns Reveal Risk in Private Banking?* (EFG Bank)

By grounding the competition in genuine institutional challenges, RiskON encourages rigorous applied research and cross-disciplinary collaboration in a way that few academic settings can replicate.

A defining feature of the RiskON hackathon is its expert jury, drawn from senior figures across the Swiss financial landscape, including former regulators and senior partners from major institutions. Prominent members include experts from PwC Switzerland, the former Vice Chairman of FINMA, and representatives from the Swiss Risk Association, granting the evaluation process significant credibility and professionalism. The jury serves a dual function: evaluating proposed solutions against demanding criteria — covering strategic relevance, technical robustness, and real-world applicability — while also offering participants direct, constructive feedback throughout the process. This combination of rigorous assessment and active mentorship ensures that winning concepts reflect both academic ambition and practical credibility

The overarching purpose of this report is to offer an academically grounded perspective on the solutions developed by the winning and runner-up teams across all three RiskON 2025 challenges. Each team brought a distinct set of methodological approaches to their respective problem, deploying techniques spanning Natural Lan-

guage Processing (NLP), Large Language Models (LLMs), Graph Neural Networks (GNNs), and anomaly detection frameworks to tackle documentation quality assurance, identity matching, and fraud detection in private banking. Collectively, these solutions demonstrate how targeted AI applications can meaningfully augment compliance workflows, sharpen fraud prevention, and enhance documentation standards while also preserving the essential role of human judgment and oversight.

The report is organized into three chapters, one dedicated to each challenge. Each chapter presents the industry-defined problem, critically evaluates the solutions put forward by competing teams, and discusses the broader implications, limitations, and potential applicability of each approach within the contemporary risk management landscape.

Chapter 1

Quality Assurance for Client Contact Notes — How can AI help?

1.1 Introduction

Client Contact Notes (CCNs) serve as an essential part of compliance in international banking. They provide documented proof of client interactions and ensure compliance with regulatory requirements, such as reverse solicitation and cross border transaction policies [Tosino et al., 2025, International Bar Association, 2024]. CCNs capture key details from conversations between relationship managers (RMs) and clients while also building the foundation of audit trails and reports. However, the traditional approach to CCN documentation presents significant challenges for financial institutions operating in multiple regulatory environments [Tosino et al., 2025, Kühne et al., 2025].

The current process for creating and reviewing CCNs has three major flaws. First, many CCNs fail to adequately address the fundamental "five Ws" (Who, What, Why, Where, and When) required for regulatory compliance [Tosino et al., 2025, Kühne et al., 2025]. This incompleteness creates gaps in documentation and may expose financial institutions to regulatory examinations and potential penalties. Second, manually creating and ensuring the quality of CCNs require more time and resources, with relationship managers often struggling to balance client service demands with administrative documentation requirements [Tosino et al., 2025, Kühne et al., 2025]. The effort required to manually review thousands of CCNs monthly places a significant burden on compliance teams. Estimates suggest that financial institutions process up to 30,000 CCNs per month in popular regions [Kühne et al., 2025]. Third, quality control processes often rely on random sampling rather than thorough or detailed review, meaning that some of the potential problems may es-

cape review entirely. These documentation challenges are further complicated by the need to operate across multiple languages and countries, with separate regulatory requirements [Tosino et al., 2025, Swiss Financial Innovation Desk, 2024].

These limitations have significant consequences. For example, incomplete or inadequate CCNs can lead to regulatory violations, particularly concerning reverse solicitation documentation where clear evidence of client initiative is legally required. The distinction between client initiated transactions and bank-initiated solicitation is crucial and must be clearly documented to avoid potential license violations and criminal liability in certain countries [International Bar Association, 2024]. Additionally, poor documentation quality results in an increase in banking operational risk, complicates dispute resolution, and undermines an institution's ability to demonstrate compliance during regulatory examinations [Tosino et al., 2025].

The 2025 RiskON challenge proposed by Bank Julius Bär, which addressed quality assurance and compliance with regulatory requirements, called for innovative technological solutions to resolve these issues. The challenge aimed to improve efficiency, effectiveness, and compliance with the standards for creating CCN and review processes by applying advanced technologies such as Natural Language Processing (NLP), Large Language Models (LLMs), and AI text analytics. The goal is to transform how relationship managers document client interactions and how compliance teams assure quality at scale.

1.2 Literature

In recent years, the application of artificial intelligence (AI) in banking documentation and compliance has received significant attention. This is due to the growing complexity of financial regulations and the increasing number of transactions that require oversight [European Banking Authority, 2023, Journal WJAETS, 2025]. Financial institutions are under increasing pressure to maintain comprehensive audit trails while managing costs and ensuring accuracy and compliance across diverse regulatory frameworks [Logwise, 2025, Banking Journal, 2025].

Natural Language Processing (NLP) is a technology that has become particularly relevant for quality assurance in the financial industry. NLP techniques enable systems to automatically analyze texts, identify patterns, extract key information, and evaluate completeness independently. Chen et al. [2021] show that NLP systems can effectively parse regulatory texts and organizational compliance policies, identifying specific requirements and obligations embedded within complex documents [International Journal of Scientific Research & Analysis, 2025]. This capability naturally extends to analyzing client contact notes, which require similar extraction and validation tasks [Mercity.ai, 2023].

Large Language Models represent a significant advancement in NLP capabilities. LLMs offer pre-trained systems that can generalize across various text analysis tasks with minimal customization. Unlike traditional machine learning models, which require extensive training data specific to a single domain, LLMs such as GPT-4 and Claude can adapt to new use cases, as they are trained on a wide variety of text sources [IBM Think, 2024]. These models have demonstrated proficiency in tasks relevant to CCN quality assurance, including text classification, entity recognition, policy identification, and the extraction of the corresponding regulatory obligation [Journal WJAETS, 2025, International Journal of Scientific Research & Analysis, 2025]. Research indicates that transformer-architecture-based models can capture complex semantic structures that simpler models miss, which enables a more accurate assessment of document completeness, quality assurance and compliance [University of Zurich Department of Finance, 2024].

The financial industry has begun using AI to automate various compliance tasks, establishing a standard for CCN quality assurance. JP Morgan Chase has implemented machine learning systems to analyze transaction data and applicant information, which has significantly reduced processing time while improving accuracy [Digital Defynd, 2025]. Similarly, Credit Suisse, which has since merged with UBS, has implemented AI techniques for mortgage underwriting that process complex applicant data more quickly and accurately than manual methods [Digital Defynd, 2025]. These applications show that AI is capable of processing large volumes of documents while maintaining quality standards and adhering to regulatory requirements.

Recent studies have shown that using LLMs for regulatory compliance automation has promise. A framework developed for real-time regulatory intelligence demonstrated that LLMs can continuously monitor regulatory sources, complex requirements, and implement compliant controls with significant improvements in interpretation accuracy and reduced time compared to traditional approaches [Journal WJAETS, 2025]. The layered architecture of this framework provides a model that can be applied to CCN quality assurance systems, which includes data ingestion, regulatory interpretation, policy management, and human oversight.

Multilingual document processing is another area in which AI technologies offer significant benefits for international banking operations. For example, financial institutions operating across borders must manage transaction documentation in multiple languages while ensuring consistent compliance standards. Advanced models have demonstrated the ability to process documents across languages without requiring separate training for each linguistic context [University of Zurich Department of Finance, 2024]. This multilingual capacity is crucial for CCN quality assurance as relationship managers may document client interactions in various languages depending on client preferences and local market practices.

Recent research on AI implementation in regulated environments has addressed the difficulty of balancing automation with human oversight [EdStellar, 2025, IBM Think, 2024, Journal WJAETS, 2025]. Although AI systems can improve efficiency and reduce errors, financial institutions must maintain appropriate human involvement to ensure accountability and manage ambiguous or borderline cases that automated systems may not handle correctly. Research suggests that hybrid approaches, in which AI performs the initial analysis and flags potential issues for human review, achieve optimal results by combining machine efficiency with human judgment [Journal WJAETS, 2025, University of Zurich Department of Finance, 2024].

1.3 Proposed Approach

To address the range of documentation and compliance challenges in Client Contact Notes (CCNs), multiple student teams proposed AI-enabled solutions targeting different stages of the CCN lifecycle. These approaches span real-time note creation, conversational guidance during documentation, and comprehensive quality assurance review. The solutions leverage advanced technologies such as Natural Language Processing (NLP), Large Language Models (LLMs), and speech-to-text transcription to transform how relationship managers document client interactions and how compliance teams ensure quality at scale. This section presents the proposed solutions and outlines their underlying methodologies, highlighting how each approach addresses a specific aspect of CCN quality management—from reducing the administrative burden on relationship managers to enabling systematic compliance monitoring across tens of thousands of monthly interactions.

1.3.1 Team 1: The AI Buddy

The AI Buddy solution employs a three-stage architecture designed to transform client interaction recordings into complete, compliance-ready Client Contact Notes. The system enables efficient documentation while maintaining regulatory standards and reducing the manual burden on relationship managers.

The solution leverages speech-to-text transcription technology to automatically process audio, text, or transcript inputs from client meetings. Transcription is performed using OpenAI Whisper, which demonstrates approximately 95% accuracy on high-quality recordings and 85–90% accuracy under noisier conditions. This multi-channel approach accommodates the multilingual conversations characteristic of Swiss private banking environments, with typical processing times of two to three minutes for a 30-minute client interaction.

Entity extraction and structured mapping are then performed using a Large Language Model that identifies key compliance-relevant information—including client

names, the discussed financial products, the involved advisors, the covered topics, and documented outcomes—and organizes this information according to Bank Julius Bär's standardized CCN template. Prompt engineering ensures consistent capture of compliance-critical elements such as client instructions, risk disclosures, and solicitation origin, without requiring model retraining for each use case.

To balance documentation quality with operational efficiency, the implementation incorporates a dual-validation mechanism. A rule-based compliance filter scans generated notes for potential regulatory issues, including references to unauthorized communication channels or insufficient documentation of client-initiated requests. A completeness verification module evaluates coverage of the Five Ws (Who, What, Why, Where, When), ensuring that all required contextual elements are present. When gaps or ambiguities are detected, the system generates targeted clarifying questions directed to the relationship manager, such as "Did the client request this product, or did you introduce it?". This iterative feedback process not only enhances final note quality but also serves an educational function, reinforcing compliance with the best practices among front staff.

1.3.2 Team 2: CCN Guardian

The development and evaluation of CCN Guardian followed a structured methodology combining workflow analysis, regulatory requirements mapping, conversational system design, AI model integration, and human-in-the-loop validation. This approach ensures alignment with both regulatory expectations and real-world RM behaviour.

The first step analysed the existing end-to-end CCN creation process through internal documentation and interviews with Bank Julius Bär staff. The analysis confirmed that CCNs are typically written after multiple client interactions, often at the end of the day. This temporal distance leads to reduced accuracy, incomplete reconstruction of decision-making processes, and missing cross-border context. The analysis identified a clear documentation gap between the client interaction and CCN entry, which became the primary intervention point for the proposed solution.

Next, regulatory obligations and internal CCN guidelines were mapped into a structured requirements matrix, including record-keeping standards, reverse solicitation rules, cross-border documentation requirements, and suitability obligations. From this mapping, functional documentation requirements were derived, including systematic capture of the Five Ws (Who, What, Why, Where, When), explicit documentation of client motivation and decision rationale, differentiation between client-initiated and RM-initiated interactions, recording of communication channels and client location, and identification of cross border triggers and solicitation indicators.

Rather than relying on form-based input, the methodology centres on a guided conversational interaction journey. RMs initiate contact with the AI agent immediately after a client interaction via a dedicated internal phone number or secure chat interface. The conversation combines fixed compliance questions with dynamic follow-up generated prompts, based on detected entities, ambiguities, or missing elements. This hybrid approach allows the system to extract complete information without overwhelming the RM or requiring explicit compliance knowledge.

To meet strict data privacy and regulatory requirements, the solution is designed for deployment within the bank's controlled infrastructure. Two deployment pathways are supported: an open-source LLM deployment enabling fully on-premise inference, and an internal bank LLM deployment leveraging institution-specific language models. Both pathways incorporate a governance layer including audit logging, rule-based safeguards for high-risk interpretations, and ongoing model performance reviews to ensure regulatory transparency.

The AI generates draft Client Contact Notes (CCNs) and delivers them to the relationship manager via the CRM backend for review and approval. CCNs are stored only after RM validation, maintaining human accountability and professional judgment. The RM engages the AI agent immediately after a client meeting or call, leveraging fresh memory while avoiding disruption to the actual interaction. The AI guides the RM through a structured dialogue aligned with internal standards, systematically capturing participants, topics and decisions, rationale, location details, and timing. Targeted questions address solicitation origin, cross-border relevance, and decision context.

The interaction is adaptive rather than checklist-driven. The agent identifies gaps and asks follow-up questions only when necessary. RMs respond in natural language while the AI semantically extracts key entities, rationale, and compliance signals, reducing cognitive effort. The AI then synthesizes the collected information into a complete CCN draft featuring consistent wording, full Five Ws coverage, explicit decision flow, and compliance with type-specific requirements.

The draft appears in the CRM environment for RM review and confirmation. While voice is the primary interaction mode, the system also supports text-based and hybrid input. Human oversight is preserved throughout: the AI assists with structuring and drafting but never autonomously finalizes or stores CCNs.

1.3.3 Team 3: Quality Assurance for Client Contact Notes

The development of the proposed solution followed a structured methodology combining problem analysis, data understanding, AI model experimentation, and prototype design.

First, the team conducted an in-depth assessment of the current CCN process

at Bank Julius Bär, mapping how relationship managers produce contact notes and how Subject-Matter Experts (SMEs) and Compliance units review them. This process analysis identified key pain points, including missing information, lack of standardisation, and inefficient supervision, quantified through interviews and operational data such as approximately 30,000 CCNs per month in the Booking Center Switzerland of Bank Julius Bär.

Second, an evaluation framework based on the "Five Ws" (Who, What, Why, Where, When) was defined. This well-established structure for assessing completeness and context in financial documentation forms the backbone of both training prompts and model evaluation criteria. Each CCN is scored along these five dimensions.

Third, the AI modelling phase tested several large language models using few-shot learning to classify and evaluate CCNs. The models included GPT-4o for its accuracy and reasoning capabilities, and Llama 3.1 (8B) for potential on-premise deployment. Both models were compared in terms of coherence and alignment with internal compliance guidelines. The model pipeline was implemented in Python, combining Flask for backend services, JavaScript for interactive frontend components, and pandas together with OpenAI API connectors for data handling. Input data consisted of anonymised CCNs exported from Excel sheets and pre-processed to remove personal identifiers and normalise text length.

The resulting prototype integrates two functional components built on the same AI core. The RM-facing assistant supports relationship managers in drafting high-quality CCNs through a chatbot-like interaction that prompts for missing information according to the Five Ws and generates a structured draft. The SME-facing dashboard automates quality assurance by evaluating submitted CCNs, calculating completeness scores, highlighting less robust sections or answers in need of improvement, and ranking notes by risk level to enable targeted review.

Finally, a human-in-the-loop evaluation strategy was adopted. SMEs and RMs tested the prototype iteratively, comparing AI-generated evaluations with manual assessments. Feedback loops allowed continuous model refinement and ensured alignment with operational needs, regulatory expectations, and user experience standards, while preserving human oversight at all critical decision points.

1.4 Discussion

The 2025 RiskON Challenge yielded three distinct AI-powered approaches to address Client Contact Note quality assurance in private banking. Each solution targets different aspects of the CCN creation and review process while sharing common foundations in natural language processing and large language models. This unified discussion assesses the strengths, limitations, and complementary aspects of all three

proposals.

The AI Buddy demonstrates strong potential for transforming CCN creation through real-time assistance. Testing on simulated client interactions showed robust transcription accuracy using Whisper (approximately 95% on clear recordings and 85-90% on noisier files) and effective LLM-generated draft notes. The compliance module successfully identified missing content or sections, where identified risk factors required attention. While a false-positive rate of 15-20% is high, it is acceptable for early-stage testing.

Pilot relationship managers reported significant time savings of 40-50% compared to manual CCN entry and valued the educational feedback provided by the system. The solution's ability to handle multilingual conversations with minimal accuracy degradation is particularly valuable for Swiss banking operations. However, pilot users stressed the importance of proper training to prevent over-reliance on AI-generated suggestions, highlighting the critical need for human oversight in maintaining documentation integrity.

CCN Guardian takes a fundamentally different approach by embedding quality assurance directly into the documentation workflow rather than relying on post-hoc review. By positioning the AI interaction immediately after client meetings, the solution addresses the most vulnerable moment in documentation quality, the gap caused by delayed recall and competing priorities.

The conversational format demonstrates several key advantages. It reduces cognitive burden by allowing relationship managers to focus on describing events rather than recalling compliance requirements, improving usability and reducing adoption resistance. Targeted prompts ensure consistent documentation of solicitation origin, client location, and decision rationale, making for a robust use case vis-a-vis the regulators.

The approach transforms traditional reactive quality controls into preventive quality assurance, reducing the need for downstream remediation and follow-up. The RM remains responsible for reviewing and approving the final CCN, ensuring alignment with supervisory expectations and professional judgment. However, potential limitations include speech recognition accuracy, model interpretability, evolving regulatory requirements, and user adoption challenges—all requiring robust governance, continuous monitoring, and comprehensive training.

The third solution offers a comprehensive framework combining an RM Assistant for front-end CCN creation and an SME Dashboard for back-end quality monitoring and compliance review. This dual approach addresses both creation and oversight challenges simultaneously.

The methodology demonstrates rigorous development, incorporating workflow analysis, regulatory requirements mapping using the Five Ws framework, and iterative human-in-the-loop evaluation. Testing multiple large language models (GPT-4o

and Llama 3.1 8B) provides flexibility for different deployment scenarios, with on-premise options addressing data privacy concerns.

From an operational perspective, the system delivers measurable benefits. Business impact simulations for the Booking Center Switzerland suggest potential savings of approximately 1,950 hours per month and CHF 2.3 million annually. The transparent scoring framework strengthens auditability and regulatory confidence. The literature review confirms alignment with industry practices, noting that regulators and industry leaders such as ESMA, PwC, and the American Bankers Association recognize LLMs as valuable compliance tools when supported by robust governance and human oversight.

All three solutions align with the broader literature on AI implementation in regulated environments, which emphasizes balancing automation with human oversight. Research suggests that hybrid approaches—where AI performs initial analysis and flags potential issues for human review—achieve optimal results by combining machine efficiency with human judgment.

The solutions demonstrate complementary strengths. The AI Buddy excels in multilingual transcription and real-time assistance; CCN Guardian focuses on preventive quality assurance through structured post-interaction capture; and the dual-module solution provides comprehensive support across both creation and review workflows with quantified business impact.

Common challenges across all solutions include maintaining appropriate human involvement to ensure accountability, managing ambigious cases that automated systems may not handle correctly, addressing data privacy and model explainability requirements, and ensuring continuous alignment with evolving regulatory requirements.

Future development should prioritize several key areas. Data privacy and on-premise deployment capabilities are essential for handling sensitive client information. Extending frameworks to additional communication channels such as emails and call transcripts would broaden applicability. Continuous model refinement through feedback loops can improve accuracy and reduce false-positive rates.

Bibliography

Banking Journal. Preparing for 2025: Navigating compliance in a time of change. https://bankingjournal.aba.com/2025/01/preparing-for-2025-navigating-compliance-in-a-time-of-change/, 2025. Accessed: 2025-11-09.

Y. Chen, X. Liu, and W. Zhang. Automated Interpretation of Financial Regulations Using Natural Language Processing. *International Journal of Scientific Research & Analysis*, 10(4):156–171, 2021.

Digital Defynd. AI in Banking: 20 Case Studies. https://www.mercity.ai/blog-post/nlp-and-llm-in-accounting,https://digitaldefynd.com/IQ/ai-in-banking-case-studies/, 2025. Accessed: 2025-11-09.

EdStellar. AI in Banking in 2025: 5 Main Uses & Tools Explained. https://www.edstellar.com/blog/ai-in-banking, 2025. Accessed: 2025-11-09.

European Banking Authority. Special topic: Artificial intelligence. https://www.eba.europa.eu/publications-and-media/publications/special-topic-artificial-intelligence, 2023. Accessed: 2025-11-09.

IBM Think. Integrating Generative AI into the Financial Regulatory Framework. https://www.mercity.ai/blog-post/nlp-and-llm-in-accounting,https://www.ibm.com/think/insights/maximizing-compliance-integrating-gen-ai-into-the-financial-regulatory-framework, 2024. Accessed: 2025-11-09.

International Bar Association. Cross-border lending to Italian borrowers by non-authorised non-EU banks and reverse solicitation. https://www.ibanet.org/cross-border-lending-to-italian-borrowers-by-non-authorised-non-eu-banks-and-reverse-solicitation, 2024. Accessed: 2025-11-09.

International Journal of Scientific Research & Analysis. Automated Interpretation of Financial Regulations Using NLP. https://journalijsra.com/sites/default/files/fulltext_pdf/IJSRA-2025-1580.pdf, 2025. Accessed: 2025-11-09.

Journal WJAETS. Real-time regulatory intelligence framework: LLM-powered compliance automation for financial services. https://journalwjaets.com/content/real-time-regulatory-intelligence-framework-llm-powered-compliance-automation-financial, 2025. Accessed: 2025-11-09.

S. Kühne, C. Wille, C. Lin, and K. Shanmugam. Quality Assurance for Client Contact Notes – How can AI help? In *RiskOn Hackathon Presentation, Team 1*, 2025.

Logwise. 5 Bank Compliance Policies for 2025 with compliance procedures. https://www.logwise.com/insights/5-bank-compliance-policies-for-2025-with-compliance-procedures/, 2025. Accessed: 2025-11-09.

Mercity.ai. NLP and LLM Applications in Accounting. https://www.mercity.ai/blog-post/nlp-and-llm-in-accounting, 2023. Accessed: 2025-11-09.

Swiss Financial Innovation Desk. Cross-Sector AI Insights for Financial Innovation. https://find.swiss/find-library/articles/cross-sector-ai-insights-for-financial-innovation, 2024. Accessed: 2025-11-09.

A. Tosino, A. Obilisetty, E. Oyelana, and F. Michler. CCN Guardian: AI assistant for RMs. In *RiskOn Hackathon Presentation, Team 2*, 2025.

University of Zurich Department of Finance. Innovation in Risk Management: Three AI-Enabled Solutions for Swiss Banking. Research paper, University of Zurich, 2024.

Chapter 2

KYC vs. Adverse Media - Can AI match identities across sources?

2.1 Introduction

Financial institutions face increasing regulatory pressure to strengthen compliance frameworks, particularly in customer due diligence and anti–money laundering (AML) activities. Inadequate screening mechanisms have resulted in substantial financial penalties and reputational damage. For example, HSBC was fined $1.9 billion in 2012 for insufficient money laundering controls and weak screening of high risk affiliates [BBC News, 2012].

A key area of vulnerability lies in Adverse Media Screening (AMS), which represents the process of analyzing news and third-party data sources to identify potentially negative information about individuals or organizations[1]. As such, AMS plays a vital role in Know Your Customer (KYC) and AML processes, but is challenged by the vast, unstructured, and often unreliable nature of online information. Traditional rule-based systems struggle to effectively manage data volume and ambiguity, leading to false positives, missed alerts, and inefficient manual reviews.

To address these limitations, financial institutions and researchers increasingly apply artificial intelligence (AI) and natural language processing (NLP) techniques to automate and enhance AMS. These approaches promise greater efficiency and accuracy in detecting relevant adverse information, but also introduce new challenges in data retrieval, text preprocessing, and analytical modeling.

The RiskON Challenge, an initiative by Zürcher Kantonalbank, addresses the technical complexities of identity matching within Adverse Media Screening (AMS). Specifically, the challenge explores the efficacy of AI in verifying whether adverse mentions are correctly attributed to a target entity—a process critical for ensuring

[1] "Adverse Media screening involves the introspection of news and other third-party data sources for potential indicators of negative news associated with an entity (person or company)." [Khandpur et al., 2021]

regulatory compliance and preventing erroneous associations. This paper provides a comprehensive review of the methodologies submitted during the challenge, contextualizing them within the current state of the AMS field. Furthermore, it offers strategic insights into improving the reliability, scalability, and transparency of automated screening frameworks.

2.2 Literature Review

Given the interdisciplinary nature of AMS and the limited academic attention it has received [Juliandri et al., 2024], this review adopts a process-oriented perspective reflecting how AMS systems are implemented in practice. The AMS workflow is strcutured in retrieval, preprocessing, and analytical modeling. Although data retrieval has been understudied, it is the foundation of any AMS system. Most studies rely on the scraping of publicly available information from news outlets and social media platforms [Juliandri et al., 2024, Khandpur et al., 2021, Roy, 2024]. However, such sources vary widely in quality and credibility and often include misinformation or bias. Uncontrolled web crawling also produces massive datasets containing millions of entries [Marko and Ries, 2023]. To mitigate these issues, Marko and Ries [2023] proposed a centralized data collection framework, where organizations can access a shared repository of verified and securely managed information. This system reduced storage needs to about one third of the original data volume and could scrape roughly 15,000 articles per day, enabling near continuous monitoring. However, the authors note that ensuring accuracy and verification of the data remains a major challenge.

After retrieval, data preprocessing is essential for preparing text for machine learning. Common steps include cleaning (removing punctuation, special characters, and capitalization inconsistencies) and tokenization to create structured text representations [Juliandri et al., 2024]. Studies have explored various preprocessing strategies. Khandpur et al. [2021], for example, compared traditional methods such as bag-of-words and sense2vec with manual tagging but found no significant performance improvements. Similarly, Juliandri et al. [2024] tested the approaches TF-IDF, Tokenizer, and CountVectorizer, concluding that while tokenizers sometimes yield high accuracy, they are prone to overfitting; CountVectorizer was shown to be more robust overall. Increasing the size of the dataset did not consistently improve the accuracy of the model. In contrast, research on model selection shows greater consensus. Most studies favor XGBoost as the preferred algorithm [Juliandri et al., 2024, Khandpur et al., 2021]. Although neural networks are theoretically capable of approximating any continuous function, their application in AMS has yielded mixed results [Lu and Lu, 2020]. Sachan et al. [2019] reported improvements in text classification based on LSTM, but noted the limitations of older architectures.

More recent Transformer models, such as XLNet [Yang et al., 2019] and optimized versions of BERT [Sun et al., 2019], achieved accuracies up to 85.4%, but often come at high computational cost. Consequently, many researchers still prefer to increase algorithms for their efficiency, interpretability, and reliability. Comparative studies consistently show that XGBoost outperforms alternatives [Anwar et al., 2021, Khandpur et al., 2021, Sinaga and Agustian, 2022]. While other gradient boosting frameworks like LightGBM or CatBoost may offer favorable trade offs, they remain underexplored in AMS contexts.

The performance of the model in AMS is typically evaluated using the F1 score [Juliandri et al., 2024, Khandpur et al., 2021, Roy, 2024], often supplemented by accuracy measures. In all studies, XGBoost achieves accuracy rates between 78% and 95%, with F1 scores around 0.93, thus substantially outperforming models such as SVM (F1 $\approx$ 0.81) [Juliandri et al., 2024, Khandpur et al., 2021, Roy, 2024]. XGBoost also exceeds CNNs, which require much more computational resources during training [Khandpur et al., 2021].

In general, literature indicates significant progress in the modeling and evaluation of AMS. However, fundamental challenges persist in data quality, preprocessing modernization, and information verification. Addressing these areas will be critical for the future development of AMS systems.

2.3 Proposed implementation strategies

Of the multiple AI-enabled solutions proposed to address adverse media screening-related challenges in private banking, a common three-stage pipeline structure emerged across student teams. This pipeline encompasses: 1. Customer on-boarding and entity verification, 2. Continuous risk assessment and monitoring, and 3. Real-time fraud detection and alert generation.

While this integrated framework serves as the primary focus in the following methodology section, a comparative overview of the complementary solutions proposed by other teams—including targeted approaches for transaction monitoring, AML detection, and regulatory compliance automation presented below to illustrate the breadth of AI applications in financial crime prevention.

2.3.1 Team 1

Team 1 developed a client identification and matching system designed to link individuals mentioned in media articles to existing client records within the organization's database. The solution enables timely risk assessment and compliance monitoring by accurately identifying whether a person referenced in adverse media corresponds to a known client.

The system leverages Named Entity Recognition (NER) to automatically extract mentions of specific individuals from unstructured article text. Client matching is then performed using a multi-attribute comparison framework that evaluates four key identifying fields: Name, Age, City, and Country. This multi-dimensional approach reduces the likelihood of false matches that might arise from name-only comparisons, particularly in cases involving common names or incomplete information.

To balance detection accuracy with operational efficiency, the implementation employs a similarity threshold set at 0.5, calibrated specifically to maintain false positives below 5%. This threshold setting reduces unnecessary alerts and manual review workload while prioritizing high-confidence matches that require analyst attention.

The robustness and performance of the classification model are validated using Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) metrics. These provide quantitative measures of the system's ability to distinguish true client matches from non-matches, supporting ongoing model tuning and quality assurance.

2.3.2 Team 2

Team 2 developed a client identification solution that leverages Large Language Models (LLMs) to extract individual names from news articles and match them against customer databases. The approach focuses on flexible name matching and quality control mechanisms to ensure accurate client identification while managing the inherent uncertainties of automated matching systems.

The system employs a general-purpose LLM for name extraction from unstructured news content, utilizing the model's natural language understanding capabilities to identify person or entities within diverse article formats and writing styles. To enhance extraction reliability, the team implemented custom prompting strategies specifically designed to reduce LLM hallucinations—a common challenge where models generate plausible but factually incorrect information.

Client matching is performed using fuzzy and phonetic matching algorithms, which accommodate variations in name spelling, formatting, and transliteration. This approach is particularly effective in handling names that may appear differently across sources due to cultural naming conventions, typos, or alternative romanizations.

The implementation includes a structured pipeline for handling cases where multiple customer records exhibit high similarity to an extracted name. Rather than forcing a single match decision, the system flags these ambiguous scenarios and incorporates a manual review step, allowing human analysts to apply contextual

judgment and domain expertise when automated confidence is insufficient.

This combination of advanced LLM-based extraction, flexible matching algorithms, and human-in-the-loop validation creates a balanced approach that leverages automation while maintaining quality control through strategic manual intervention points.

2.3.3 Team 3

Team 3 developed a multi-stage client identification pipeline that combines intelligent article filtering, sequential data extraction, and LLM-based refinement to accurately match individuals mentioned in adverse media to customer records. The approach emphasizes efficiency and precision through progressive filtering and enrichment steps.

The system begins with an initial filtering mechanism that screens out articles not containing adverse media content, reducing the volume of data requiring detailed analysis and focusing computational resources on relevant risk-related sources. This preliminary filter improves overall pipeline efficiency by eliminating non-critical content early in the process.

Name extraction is performed using a general-purpose LLM capable of identifying full names from unstructured article text across various formats and contexts. These extracted names serve as the basis for preliminary matching against the customer database, establishing an initial set of potential client connections.

Following preliminary matches, the system performs secondary extraction to gather additional customer-related attributes, including age, city, country of residence, and nationalities. This enrichment step provides contextual data points that support more confident matching decisions and help disambiguate between similar names.

The final stage employs a separate LLM in a patching role, synthesizing information from both the extracted article data and preliminary matches to determine the final set of matched clients. This two-LLM architecture separates initial extraction from final decision-making, allowing each model to be optimized for its specific task while providing an additional layer of validation before presenting results for review.

2.4 Methodology

This paper primarily examines the comprehensive risk management framework proposed by team 4, which demonstrated the most promising results during initial validation. The proposed approach creates an integrated customer-centric risk management system by:

- Leveraging advanced AI technologies including Large Language Models (LLMs)

and machine learning algorithms to automate Know Your Customer (KYC) verification and extract relevant customer information from multiple data sources,

- Implementing specialized predictive analytics and behavioral pattern recognition models to continuously monitor customer activities and flag potentially suspicious transactions or high-risk behaviors,
- Deploying real-time fraud detection mechanisms that combine anomaly detection, fuzzy matching techniques, and adaptive learning capabilities to identify fraudulent activities across the customer lifecycle.

2.4.1 Data

To comply with Know Your Customer (KYC) and AntiMoney Laundering (AML) regulations, financial institutions must keep extensive structured records of their clients. Information such as names, date of birth, nationalities, and place of residence is one of the key identifiers.

At the same time, these institutions must monitor unstructured external data sources, such as negative media from news articles, for potential risks associated with their clients. The methodology was developed and tested using two datasets that simulate this scenario:

- News articles dataset: Training and testing set of 10'000 and 300 news articles, respectively, with each article containing a title, name of source, date, and the article content in text format (without images, hyperlinks, or text formatting).
- Client dataset: Client information of 2+ million synthetic clients, including names, date of birth, nationalities, and place of residency.

The core methodological challenge arises directly from the disparity between these two data sources. The task is to match people identified in the unstructured news data with the structured client list. This process is complicated by the inherent ambiguity of the unstructured text, which includes spelling variations (e.g., Müller, Mueller), transliteration differences (e.g., spelling of Natalia in latin and cyrillic alphabets), and the high frequency of common names (e.g., Peter Müller in the Swiss market). In addition, matching persons by age has certain ambiguity, as one can only obtain an age range of the extracted person given the person's mentioned age and article publication date rather than a precise date of birth.

2.4.2 Extraction Step

The first stage of the pipeline addresses the challenge of processing large volumes of unstructured news text. GPT 4.1 mini, an OpenAI general purpose LLM, was chosen as the AI agent for the task of extracting people from news articles. This model was chosen for its balance between speed, relative intelligence, and token

cost. Using specific prompts, the model was instructed to read each news article and extract the key information on all mentioned individuals (names, age, date of birth, nationalities and place of residency) along with any relevant contextual details, such as their role or associated actions described in the article. Through the extraction of the person's age (if mentioned) and given the article's publication date, the extraction pipeline makes use of a method to define the age range for each extracted person. Parallel processing was implemented in the extraction step to improve the speed of persons extraction from news articles. This approach effectively solves the problem of handling articles that mention multiple people, ensuring that no potential subjects are missed. In the training phase, out of the 10'000 news articles available, 22'674 unique individuals were extracted.

2.4.3 Filtration Step

Following extraction, a critical filtration step is applied to distinguish high risk individuals from less relevant mentioned persons (e.g., journalists, by-passers, or academics), thereby reducing the number of false positives. This stage utilizes a hybrid approach using FlashText[2] for quickly scanning articles from a list of manually created "risk keywords", such as "fraud", "laundering" and "accused". This approach allows for the flagging of potential links to criminal or high risk activities. In the testing phase, this step successfully filtered out 42% of extracted persons (9'530 individuals), while retaining 58% (13'114 individuals) who were contextually linked to potential risks (e.g., "accused of credit card fraud").

2.4.4 Matching Step

The final stage involves matching the filtered, high risk individuals against the 2+ million synthetic client records. This step is designed to overcome common data matching problems like spelling variations, common names, and transliteration differences. A two stage matching process was implemented. The core of this process utilizes RapidFuzz[3]. To enhance accuracy, this is supplemented by a Phonetics tool, which compares names based on their sound rather than just their spelling. The final output generates a confidence score based on a weighted match of multiple identifiers, including Name, Phonetics, Age, Location, and Nationality. A match is identified when a certain confidence score threshold is reached.

[2]FlashText is a Python library designed for fast keyword searching and replacement in text. It is specifically optimized for large keyword lists for fast finding or replacement of keywords [Singh, 2017]

[3]RapidFuzz is library for high speed fuzzy string matching on names [Ye et al., 2021]

2.4.5 Results

In the training phase, the pipeline demonstrated promising results. The filtration step achieved a false negative rate of 0% on the training data, successfully identifying all known high risk individuals in the set. In the final matching phase, the system correctly identified 81.8% of all true matches between the news articles and the client list. This was accompanied by 303 false positives. It is important to note that these figures, while encouraging, are based on preliminary testing and may not precisely represent the true predictive capabilities of the solution on a production scale dataset.

2.5 Discussion

The results of this study demonstrate that AI supported pipelines hold considerable promise to enhance AMS and identity matching. Across all teams in the RiskON challenge, a common conceptual architecture emerged as follows: entity extraction, risk filtration, and then client matching. This indicates a convergence in practices that intuitively decompose the AMS problem. This convergence not only highlights the clarity of the problem structure, but also exposes shared methodological challenges that remain unresolved.

A notable finding is the strong performance of Large Language Models in the extraction stage. Team 4's implementation, which relied on GPT 4.1 mini, showed that general purpose LLMs can reliably parse unstructured text and identify individuals mentioned in news articles, even when several appear in a single document. This suggests that LLMs can replace, or at minimum substantially augment, traditional Named Entity Recognition approaches. However, the extraction process still faces ambiguity stemming from inconsistent reporting styles, varying linguistic conventions, and missing or conflicting attributes such as age or nationality. Although parallelization mitigated processing delays, extraction accuracy is ultimately bound by the model's interpretation of context—an issue that becomes increasingly prominent as datasets scale or include multilingual content.

Albeit the filtration step is effective in reducing noise, it also raises important considerations. The hybrid FlashText keyword method used by Team 4 achieved a strong reduction in irrelevant entities, yet it relies fundamentally on manually curated lists of risk indicators and is therefore sensitive to keyword selection. Although the approach achieved a false negative rate of 0% in the training set, this figure must be interpreted with caution. Real world data might be more heterogeneous than synthetic Hackathon datasets. Criminal activity is not always described with explicit keywords, and media reports may use euphemisms or indirect phrasing. Thus, while keyword methods offer speed and simplicity, they may underperform

in nuanced contexts, suggesting that future risk screening models should incorporate semantic analysis, LLM enhanced classification, or hybrid approaches to reduce reliance on lexical cues.

The matching stage underscores the central tension in AMS: balancing recall with precision. Team 4's combination of fuzzy matching and phonetic similarity represents a practical and computationally efficient approach. Its performance successfully identifying $\approx 80\%$ of true matches shows that relatively lightweight matching algorithms can deliver strong results when combined with enriched entity attributes extracted by LLMs. Still, the 303 false positives highlight the persistent difficulty in resolving identities when names are common, transliterated, or inconsistently spelled. As noted in the previous literature, variation in spelling (e.g., Müller vs. Mueller) and high frequency names remain major sources of error. The inclusion of additional structured attributes (e.g., age ranges, nationality codes, historical addresses) could meaningfully reduce ambiguity, but this would require more reliable extraction methods and more sophisticated weighting strategies in the matching score.

2.6 Limitations and Future Research

All approaches reveal a methodological limitation: the heavy dependence on textual attributes extracted from noisy, sometimes incomplete news sources. As financial institutions increasingly face multilingual, multimedia, and cross platform streams of information, AMS systems will require more robust, context sensitive models capable of integrating information beyond names and simple demographic attributes. In addition, the economic viability of these pipelines remains an open question. LLM based extraction is computationally expensive at scale, and while smaller models like GPT 4.1 mini offer a good compromise, the efficiency performance trade off will become increasingly critical as organizations expand continuous monitoring to millions of articles per day.

The findings suggest that AI can substantively improve AMS processes, particularly in entity extraction and structured matching. However, the systems tested still lack the reliability required in high stakes compliance environments. For AI supported AMS to reach operational maturity, future research must address ambiguity in unstructured text, improve entity resolution under uncertainty, and develop cost effective strategies for real time monitoring at scale. The challenge ahead is not simply to enhance accuracy but to build systems that are transparent, auditable, and robust enough to withstand regulatory scrutiny.

Bibliography

M. T. Anwar, A. E. Pratiwi, and K. Udhayana. Automatic complaints categorization using random forest and gradient boosting. *Advance Sustainable Science, Engineering and Technology*, 3(1):0210106, 2021.

BBC News. HSBC to pay $1.9 bn in US money laundering penalties. https://www.bbc.com/news/business-20673466, December 2012. Accessed: 2025-10-27.

R. Juliandri, M. E. Johan, J. Wiratama, and S. A. Sanjaya. Adverse media classification: A new era of risk management with XGBoost and gradient boosting algorithms. In *2024 5th International Conference on Big Data Analytics and Practices (IBDAP)*, pages 18–21. IEEE, 2024.

R. P. Khandpur, A. A. Nanda, M. Davis, C. Li, D. Nurmanbetov, S. Gaur, and A. Talukder. Adverse media mining for KYC and ESG compliance. *arXiv preprint arXiv:2110.11542*, 2021.

Y. Lu and J. Lu. A universal approximation theorem of deep neural networks for expressing probability distributions. *Advances in Neural Information Processing Systems*, 33:3094–3105, 2020.

R. Marko and M. Ries. Adverse media screening portal. In *2023 Communication and Information Technologies (KIT)*, pages 1–6. IEEE, 2023.

S. Roy. AI Adoption to Combat Financial Crime: Study on Natural Language Processing in Adverse Media Screening of Financial Services in English and Bangla Multilingual Interpretation. *arXiv preprint arXiv:2412.12171*, 2024.

D. S. Sachan, M. Zaheer, and R. Salakhutdinov. Revisiting LSTM networks for semi-supervised text classification via mixed objective function. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):6940–6948, 2019.

H. H. Sinaga and S. Agustian. Perbandingan Metode Decision Tree dan XGBoost untuk Klasifikasi Sentimen Vaksin Covid-19 di Twitter. *Perbandingan Metode Decision Tree dan XGBoost untuk Klasifikasi Sentimen Vaksin Covid-19 di Twitter*, 8(03):107–114, 2022.

V. Singh. Replace or retrieve keywords in documents at scale. *arXiv preprint arXiv:1711.00046*, 2017.

C. Sun, X. Qiu, Y. Xu, and X. Huang. How to fine-tune BERT for text classification? In *China National Conference on Chinese Computational Linguistics*, pages 194–206. Springer, 2019.

Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le. XLNet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, 32, 2019.

A. Ye, L. Wang, L. Zhao, J. Ke, W. Wang, and Q. Liu. RapidFuzz: Accelerating fuzzing via generative adversarial networks. *Neurocomputing*, 460:195–204, 2021.

Chapter 3

AI-powered Fraud Detection – How can patterns reveal risk in private banking?

3.1 Introduction

Fraud and financial crime have long posed significant challenges to the global financial system, predating the emergence of modern digital technologies. Historically, fraudulent activity has taken many forms, including identity misrepresentation, insider abuse, money laundering, and transaction manipulation, often exploiting information asymmetries and operational weaknesses within financial institutions. Despite extensive regulatory frameworks and control mechanisms, fraud remains persistent, adaptive, and costly, driven by both economic incentives and evolving financial infrastructures.

In recent decades, the increasing digitization of banking services has transformed the scale and complexity of fraud. High transaction volumes, real-time payment systems, global connectivity, and the proliferation of online and mobile banking channels have expanded the attack surface available to fraud actors. Traditional fraud detection mechanisms, which rely heavily on static rules and manual investigation, are increasingly strained by the velocity, variety, and volume of modern financial data.

More recently, advances in artificial intelligence have altered the fraud landscape in two distinct ways. On one hand, AI and data-driven analytics provide powerful tools for detecting anomalous behavior, uncovering hidden relationships, and automating compliance processes. On the other hand, the same technological advances have been leveraged by fraudsters to enhance deception techniques. AI-assisted fraud typologies–such as synthetic identity creation, voice cloning, and deepfake video impersonation–enable attackers to more convincingly mimic legiti-

mate customers and evade traditional verification mechanisms. These developments do not replace traditional fraud but rather augment existing schemes, increasing their scale, speed, and sophistication.

The impact of fraud extends beyond direct financial losses to include operational disruption, reputational damage, and regulatory consequences. Financial institutions must therefore address a broad spectrum of risks that span customer onboarding, ongoing account activity, and real-time interactions between clients and employees. In this context, Know Your Customer (KYC) processes establish baseline identity assurance, while continuous monitoring and interaction-level controls are required to capture evolving and situational risks.

In light of these developments, the RiskON 2025 Challenge 3, proposed by EFG Bank, explores how artificial intelligence can be leveraged to strengthen fraud detection and risk management in private banking. Rather than framing AI as a replacement for existing controls, the challenge emphasizes its role in enhancing traditional approaches through automation, relational analysis, and adaptive learning. Participants were asked to conceptualize AI-driven solutions that address both external and internal fraud risks across different stages of the customer lifecycle.

This paper synthesizes and compares several proposed approaches developed for the challenge, spanning automated KYC verification, continuous customer risk monitoring, and real-time fraud detection during live interactions. Due to the absence of operational datasets, the focus is placed on system architecture and methodological design rather than empirical evaluation. Together, the proposed solutions illustrate how AI can complement established fraud prevention strategies and support more adaptive, scalable, and transparent risk management in private banking environments.

3.2 Literature Review

Fraud detection and regulatory compliance have long been central research areas in banking and financial services. Early fraud detection systems relied primarily on statistical techniques and rule-based approaches that used predefined thresholds and expert-crafted rules to flag suspicious behavior. While such systems offer transparency and interpretability, extensive research has shown that they struggle with adaptability and often generate high false-positive rates, particularly in environments characterized by evolving fraud strategies and highly imbalanced datasets [Bolton and Hand, 2002, Ngai et al., 2011]. These limitations motivated the development of more flexible, data-driven approaches.

Machine learning techniques have increasingly been adopted to address the rigidity of traditional fraud detection systems. Supervised learning models can capture complex, non-linear relationships in transactional data and identifying subtle pat-

terns indicative of fraudulent activity. However, fraud detection presents unique challenges, including class imbalance, delayed labels, concept drift, and adversarial behavior, as fraudsters continuously modify their tactics to evade detection. Research on adaptive and drift-aware fraud detection highlights the importance of continuous model updates and incremental learning mechanisms to maintain detection performance over time [Dal Pozzolo et al., 2018, West and Bhattacharya, 2016]. These challenges are particularly relevant for Know Your Customer (KYC) processes, which increasingly require continuous monitoring rather than static, one-time verification.

A substantial body of work has explored machine learning models for operational fraud detection in banking. Cost-sensitive learning has been shown to be particularly effective in fraud contexts, where misclassifications costs are asymmetric. Sahin et al. [2013] demonstrated that cost-sensitive decision tree models significantly improve fraud detection performance on imbalanced financial datasets. Ensemble and anomaly-based approaches, including random forests and isolation forests, have also been widely applied to credit card and transaction fraud detection, highlighting their robustness to noise and evolving fraud patterns [Whitrow et al., 2009, Xuan et al., 2018, Liu et al., 2008].

Beyond transactional modeling, Natural Language Processing (NLP) has emerged as a key enabler for automating compliance and KYC workflows, driven by the growing volume of unstructured data such as adverse media reports, legal documents, regulatory filings, and corporate disclosures. Transformer-based language models, including BERT and optimized variants such as RoBERTa and FinBERT, have demonstrated strong performance in financial text classification, named-entity recognition, and sentiment analysis [Araci, 2019, Liu et al., 2019]. These models enable scalable extraction of reputational and legal risk signals, reducing reliance on manual document review.

Prior research emphasizes the importance of domain-specific language modeling in financial and legal contexts. Loughran and McDonald [2011] showed that general-purpose sentiment dictionaries frequently misclassify financial language, motivating tailored NLP approaches for compliance applications. Advances in legal NLP further demonstrate the effectiveness of neural models in extracting structured information from legal documents, supporting applications such as litigation risk assessment and regulatory monitoring [Chalkidis and Androutsopoulos, 2019].

Beyond textual analysis, relational modeling has gained prominence as financial crime increasingly manifests through networks of interconnected entities rather than isolated actors. Graph-based approaches explicitly model relationships among customers, accounts, organizations, and transactions, enabling detection of indirect exposure and coordinated behavior. Surveys of graph-based anomaly detection techniques highlight their effectiveness in identifying suspicious structures and patterns

in complex networks [Akoglu et al., 2015].

More recently, Graph Neural Networks (GNNs) have been applied to fraud and anti-money laundering detection by integrating relational structure with node-level attributes. Architectures such as Graph Convolutional Networks (GCNs) and GraphSAGE have been shown to improve detection of illicit activity by propagating risk signals across connected entities [Weber et al., 2019, Hamilton et al., 2017]. Dynamic graph models, such as EvolveGCN, further support continuous risk assessment by updating representations as new transactions and relationships emerge [Pareja et al., 2020]. These methods are particularly relevant for KYC and customer monitoring systems that must account for evolving risk exposure over time.

In parallel, the rise of AI-enabled fraud has introduced new challenges for financial institutions. Techniques such as synthetic identity creation, voice cloning, and deepfake video impersonation exploit advances in generative models to bypass traditional verification mechanisms [Verdoliva, 2020, Mirsky and Lee, 2021]. Research indicates that detecting such attacks requires analysis of behavioral, biometric, and contextual inconsistencies rather than reliance on static identity attributes alone [Jain et al., 2011, Kotiyal et al., 2024]. This has motivated the integration of biometric analysis and real-time anomaly detection as complementary layers to traditional KYC and transaction monitoring systems.

Despite the effectiveness of advanced AI models, explainability and human oversight remain critical in regulated financial environments. Black-box decision-making systems raise concerns regarding auditability, accountability, and regulatory acceptance. Research in Explainable Artificial Intelligence (XAI) highlights the importance of interpretable models and post-hoc explanation techniques to support transparency in high-stakes applications [Doshi-Velez and Kim, 2017, Arrieta et al., 2020]. Consequently, human-in-the-loop frameworks are increasingly adopted to combine automated detection with expert judgment and regulatory compliance.

Overall, existing literature demonstrates the potential of combining transactional modeling, NLP-based intelligence extraction, graph-based relational learning, adaptive algorithms, and explainability mechanisms to enhance fraud detection and KYC systems in banking. However, many studies address these components in isolation. Fewer approaches consider their integration within unified frameworks capable of supporting onboarding, continuous monitoring, and interaction-level fraud detection. The approaches explored in this work build on prior research by examining how these complementary techniques can be combined to support customer-centric risk management in private banking contexts.

3.3 Proposed implementation strategies

To address the range of fraud and compliance challenges in private banking, multiple student teams proposed AI-enabled solutions targeting different stages of the risk management lifecycle. These approaches span customer onboarding, continuous risk monitoring, and real-time interaction fraud detection. This section presents the proposed solutions and outlines their underlying methodologies, highlighting how each approach addresses a specific aspect of customer-centric risk management.

3.3.1 Team 1: AI-Powered Employee Screening Using Open-Source Intelligence

Team 1 proposed an AI-powered employee screening solution based on open-source intelligence (OSINT) to identify reputational and behavioral risks associated with individuals. The approach is primarily designed for internal risk assessment, focusing on employees or potential hires, rather than customer onboarding. The system analyzes publicly available online content, including news articles, blogs, and social media platforms, to detect adverse media, controversial narratives, or signals indicative of potential misconduct.

The solution employs web-search APIs such as Google API and SerpAPI to retrieve relevant online sources associated with an individual. Web scraping tools, including Scrapy, BeautifulSoup, and newspaper3k, are used to extract textual content from articles and online publications. Natural Language Processing (NLP) techniques are then applied to analyze tone, intent, and thematic patterns within the collected text.

Text embeddings generated using transformer-based language models such as BERT enable semantic similarity analysis and clustering of risk-related content. Sentiment and intention classification models are used to flag potentially concerning narratives, which are presented to analysts through threshold-based risk indicators. The approach relies on publicly visible information and is therefore sensitive to variations in data quality, contextual ambiguity, and the presence of misinformation, which necessitates human review.

3.3.2 Team 2: AI-Powered KYC Through Evidence-Weighted Identity Corroboration

Team 2 proposed an AI-powered Know Your Customer (KYC) solution focused on compliance-oriented identity corroboration. The approach addresses the fundamental KYC requirement of determining whether a claimed identity genuinely exists and is consistently represented across reliable data sources. To achieve this, the solution introduces a multi-tiered evidence pipeline in which data sources are assigned

different evidentiary weights.

Authoritative sources such as government registries, corporate ownership databases, court filings, and professional license records form the highest evidentiary tier. Secondary sources include reputable news outlets and academic or institutional directories, while lower-weight contextual sources include corporate websites and professional networks. Documents obtained from these sources are parsed from PDF to text to enable structured information extraction.

Entity linking techniques are applied to cluster mentions referring to the same individual across different sources. Similarity features—including name similarity, date-of-birth matches, address overlap, geospatial distance, and employer or position consistency—are computed to assess identity coherence. Each source category contributes to the overall assessment through predefined weights. Large Language Models (LLMs) are used to summarize corroborated evidence and highlight inconsistencies, such as incoherent employment histories or abnormal activity patterns. The system produces a final risk classification indicating whether the identity is strongly corroborated, requires further review, or is not sufficiently supported.

3.3.3 Team 3 (Winning Team): AI-Driven Continuous KYC Automation (KYClens)

The winning team proposed KYClens, an AI-enabled platform designed to automate and continuously update KYC risk assessments. Unlike static onboarding checks, KYClens integrates diverse external data sources—including adverse media, sanctions lists, legal records, and corporate registries—into a unified intelligence pipeline.

Transformer-based NLP models automatically extract risk-relevant entities, events, and sentiments from unstructured text, significantly reducing manual review effort. To capture complex relationships among customers, counterparties, and organizations, KYClens constructs a heterogeneous entity graph processed using Graph Neural Networks (GNNs). This enables detection of hidden associations, indirect exposure to high-risk entities, and dynamic risk propagation.

The system continuously updates customer risk scores as new information emerges, triggering alerts when relevant changes occur. A human-in-the-loop mechanism allows compliance analysts to review and correct model outputs, ensuring transparency, regulatory alignment, and continuous learning. KYClens distinguishes itself by combining scalability, adaptability, and network-aware risk detection within a single framework.

3.3.4 Team 4: Biometric Guard – AI-Powered Fraud Detection for Live Interactions

Team 4 proposed *Biometric Guard*, an individualized AI-powered fraud detection system designed to detect impersonation and AI-enabled fraud during live audio and video interactions. Unlike traditional KYC systems that focus on identity verification during onboarding, Biometric Guard targets real-time fraud risks arising during client–bank and employee–bank interactions, such as deepfake attacks, voice cloning, and synthetic video impersonation.

The system leverages biometric signals derived from facial features, voice characteristics, and behavioral patterns captured during live calls. Convolutional Neural Networks (CNNs) are used to learn personalized facial embeddings that remain robust to changes in lighting, appearance, or camera quality. Voice analysis models generate individual voiceprints based on acoustic features such as pitch, tone, and temporal dynamics, enabling detection of synthetic or manipulated speech.

In addition to biometric verification, the system incorporates knowledge-based challenges and one-time password (OTP) verification to strengthen fraud resistance. By continuously analyzing biometric consistency throughout a session, Biometric Guard can detect anomalies indicative of identity takeover or AI-generated impersonation. While this approach is not a standalone KYC solution, it complements onboarding and monitoring systems by providing an additional layer of real-time fraud protection for both clients and employees.

3.3.5 Comparative Discussion

The proposed solutions address financial crime risk at different stages of the banking lifecycle and differ primarily in scope, data modality, and timing of intervention. Some approaches focus on pre-engagement and onboarding risk assessment, others emphasize continuous monitoring of evolving customer exposure, while one targets fraud that arises during real-time interactions.

Approaches centered on open-source intelligence and evidence-weighted identity corroboration address risks associated with identity legitimacy, reputational exposure, and onboarding compliance. These methods rely largely on external data sources and structured verification logic to establish baseline risk profiles. In contrast, continuous KYC automation extends this perspective by incorporating relational and transactional data, enabling dynamic risk assessment that accounts for indirect associations and changes over time.

Real-time fraud detection during live audio and video interactions represents a distinct risk layer, addressing impersonation and AI-enabled fraud that may occur after onboarding and outside traditional KYC workflows. By focusing on biometric and behavioral signals, this category of solution complements identity-centric

systems rather than replacing them.

Taken together, the proposed approaches illustrate that effective financial crime prevention in private banking requires a layered strategy. Identity verification, continuous monitoring, and session-level fraud detection address different threat vectors and can be viewed as interoperable components within a broader risk management framework, rather than as mutually exclusive alternatives.

3.4 Methodology

Although multiple AI-based solutions were proposed during the RiskON 2025 challenge, this section focuses on the winning team's solution, *KYClens*, as a representative framework that demonstrates how several of the techniques discussed across teams can be integrated within a single system. Given the hackathon setting and the absence of access to operational banking data, the methodology is presented at a conceptual and architectural level rather than as a fully implemented system.

KYClens follows a modular, end-to-end design intended to support scalable, explainable, and continuously adaptive KYC risk assessment. The framework integrates heterogeneous data ingestion, intelligent information extraction, relational risk modeling, and human oversight into a unified workflow, enabling both customer onboarding verification and continuous risk monitoring throughout the customer lifecycle.

3.4.1 Data Ingestion and Normalization

The first stage of the KYClens pipeline focuses on the secure ingestion and normalization of heterogeneous data sources. Internal inputs include customer onboarding information, identification documents, account metadata, transaction records, device identifiers, and location logs. These are complemented by external intelligence sources such as adverse media feeds, sanctions and Politically Exposed Persons (PEP) lists, court records, corporate registries, and other compliance-relevant public datasets.

All incoming data is processed through a normalization pipeline that performs cleaning, de-duplication, tokenization, and schema alignment. Entity resolution techniques are applied to reconcile aliases, spelling variations, and incomplete records across sources. Transactional, device, and location data are temporally aligned to support longitudinal analysis. This stage produces a unified and consistent representation of customers, entities, and events, forming the foundation for downstream intelligence extraction and risk modeling.

3.4.2 NLP-Based Intelligence Extraction

Natural Language Processing (NLP) models are applied to extract structured risk intelligence from unstructured text sources. Transformer-based models such as FinBERT or RoBERTa are fine-tuned for tasks including named-entity recognition, adverse media classification, sentiment analysis, and identification of risk-indicative events such as fraud, litigation, corruption, or regulatory violations.

The NLP pipeline extracts entities including persons, organizations, locations, and legal references, along with contextual attributes such as sentiment polarity, event severity, and source credibility. Extracted information is timestamped and enriched with metadata to support temporal relevance assessment. This automated extraction enables scalable analysis of large volumes of external text while ensuring consistency and reducing manual review effort.

3.4.3 Graph Construction and Risk Propagation

To capture complex and evolving risk beyond isolated customer attributes, KYClens constructs a dynamic, heterogeneous entity graph that models relationships among individuals, organizations, accounts, transactions, devices, locations, and suspicious activity events. Nodes represent entities such as customers, corporate entities, bank accounts, devices, geographic locations, and historical risk assessments. Edges encode both direct and indirect relationships, including transaction flows, shared devices, common IP addresses, overlapping locations, corporate ownership links, and co-occurrence in adverse media or suspicious activity reports.

Transactional relationships are modeled to identify unusual money flows and indirect exposure to high-risk entities. Device and location signals are incorporated to detect identity misuse, account sharing, or coordinated behavior across multiple customers. Suspicious activity indicators and prior risk scores are attached as temporal attributes, allowing the system to maintain a historical view of risk evolution.

Graph Neural Networks (GNNs), including GraphSAGE and Graph Convolutional Network (GCN) variants, are applied to learn contextual node embeddings that integrate entity attributes, relational structure, and temporal risk signals. Risk is propagated across the graph so that newly identified suspicious activity or elevated risk associated with one entity dynamically influences connected nodes through both direct and indirect relationships. As a result, related entities sharing devices, locations, or transactional links with flagged customers are automatically re-evaluated and, where appropriate, flagged for analyst review.

The entity graph is continuously updated as new data becomes available, including incoming transactions, newly observed devices or locations, and updated adverse media signals. Risk scores are recalculated dynamically, enabling continuous KYC monitoring rather than static, point-in-time assessment. This dynamic relational

modeling supports early detection of emerging risk patterns and hidden associations that may not be observable through individual-level analysis alone.

3.4.4 Human-in-the-Loop and Explainability

Given the regulatory sensitivity of KYC decisions, KYClens integrates a human-in-the-loop framework to ensure transparency, accountability, and regulatory alignment. Compliance analysts review system-generated alerts, validate extracted intelligence, and assess flagged relationships. Analyst feedback is incorporated into model refinement through active learning mechanisms, enabling continuous improvement while maintaining human oversight.

Explainability techniques such as SHAP and LIME are integrated to provide interpretable insights into model outputs. These tools highlight influential features, entities, transactions, and relationships contributing to each risk assessment, supporting analyst decision-making and facilitating regulatory audits.

3.4.5 Deployment and Monitoring

The final stage focuses on system deployment and operational monitoring. KYClens is designed to integrate with existing banking and compliance infrastructures through secure APIs and data pipelines. An analyst-facing dashboard presents dynamic risk scores, supporting evidence, extracted entities, transaction histories, and interactive graph visualizations to facilitate efficient case review.

Once deployed, the system supports continuous monitoring of customer profiles, automatically updating risk assessments as new information emerges. Automated alerts notify analysts of significant risk changes, while comprehensive audit logs and reporting mechanisms ensure traceability and regulatory compliance. A controlled pilot phase using synthetic or anonymized datasets enables validation of performance, usability, and operational readiness prior to full-scale production deployment.

3.5 Discussion

This work examined multiple AI-based approaches to addressing fraud and risk in private banking, with an emphasis on customer-centric processes such as identity verification, continuous monitoring, and interaction-level fraud detection. The proposed solutions illustrate how different analytical techniques can be applied to distinct stages of the risk management lifecycle, reflecting the increasingly layered nature of financial crime prevention.

The KYClens framework demonstrates how automated KYC and ongoing risk assessment can be supported through the integration of NLP-based intelligence extrac-

tion, graph-based relational modeling, and dynamic risk scoring. By incorporating transactional data, device and location signals, and historical risk information, the system enables continuous reassessment of customer profiles as new data becomes available. Modeling both direct and indirect relationships allows for identification of evolving risk exposure that may not be apparent through isolated customer attributes alone. The inclusion of human oversight and explainability mechanisms reflects the regulatory requirement for transparency and accountability in automated decision-making.

The comparative analysis further highlights that not all fraud risks are confined to onboarding or periodic KYC reviews. AI-enabled fraud techniques, such as synthetic identity creation, voice cloning, and impersonation during live interactions, introduce threats that arise after initial identity verification. Solutions such as Biometric Guard address these risks by focusing on real-time behavioral and biometric analysis during audio and video interactions involving clients and employees. This type of session-level fraud detection operates independently of traditional KYC workflows and targets a different segment of the fraud landscape.

Across all approaches, several common challenges emerge. Data availability and quality remain key constraints, particularly given privacy, consent, and regulatory limitations associated with sensitive information such as biometric signals, device identifiers, and location data. In addition, the use of advanced AI models introduces requirements for interpretability, governance, and sustained human involvement to ensure responsible deployment in regulated environments. The absence of real-world datasets in this work further limits empirical validation, emphasizing the need for careful system design and future evaluation using anonymized or synthetic data.

Overall, the discussion underscores effective fraud and risk management in private banking requires multiple, complementary analytical layers rather than a single solution. KYC systems establish baseline identity and compliance, continuous monitoring frameworks track evolving exposure over time, and real-time fraud detection mechanisms address interaction-level threats, including those facilitated by AI. When considered together, these approaches illustrate how AI-driven methods can contribute to more adaptive, scalable, and transparent financial crime prevention strategies within private banking contexts.

3.6 Limitations and Ethical Considerations

Despite its potential benefits, the proposed KYClens framework faces several limitations. The primary constraint is the lack of access to real-world KYC datasets, which restrict empirical validation and performance benchmarking. Synthetic or anonymized datasets may partially address this issue but cannot fully capture operational complexity.

Ethical considerations are also critical. Automated risk assessment systems may introduce bias due to skewed data sources or historical enforcement patterns. Continuous monitoring of individuals raises privacy concerns, particularly when aggregating external intelligence. For biometric-based approaches, consent, data storage, and cross-border regulatory compliance require careful governance.

Mitigating these risks requires transparent model design, regular bias audits, strong data governance policies, and sustained human oversight.

Bibliography

L. Akoglu, H. Tong, and D. Koutra. Graph-based anomaly detection and description: A survey. *Data Mining and Knowledge Discovery*, 29(3):626–688, 2015.

D. Araci. FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*, 2019.

A. B. Arrieta et al. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges. *Information Fusion*, 58:82–115, 2020.

R. J. Bolton and D. J. Hand. Statistical fraud detection: A review. *Statistical Science*, 17(3):235–255, 2002.

I. Chalkidis and I. Androutsopoulos. Neural legal judgment prediction in English. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4317–4323, 2019.

A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi. Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy. *IEEE Transactions on Neural Networks and Learning Systems*, 2018. doi: 10.1109/TNNLS.2017.2736643.

F. Doshi-Velez and B. Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.

W. L. Hamilton, R. Ying, and J. Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pages 1024–1034, 2017.

A. K. Jain, A. Ross, and K. Nandakumar. *Introduction to biometrics*. Springer, 2011.

A. Kotiyal, L. Hussein, A. Deepak, A. Rana, Manjunatha, K. K. Dixit, and R. A. Reddy. Graph-Based Machine Learning Approaches for Fraud Detection in Financial Networks. In *Proceedings of the 2024 International Conference on Computing, Communication, and Intelligent Systems (IC3I)*, 2024. doi: 10.1109/ic3i61595.2024.10828743.

F. T. Liu, K. M. Ting, and Z.-H. Zhou. Isolation forest. In *Proceedings of the 2008 IEEE International Conference on Data Mining*, pages 413–422, 2008.

Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

T. Loughran and B. McDonald. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1):35–65, 2011.

Y. Mirsky and W. Lee. The creation and detection of deepfakes: A survey. *ACM*

Computing Surveys, 54(1):1–41, 2021.

E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3):559–569, 2011.

A. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, and C. E. Leiserson. EvolveGCN: Evolving graph convolutional networks for dynamic graphs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(4):5363–5370, 2020.

Y. Sahin, S. Bulkan, and E. Duman. A cost-sensitive decision tree approach for fraud detection. *Expert Systems with Applications*, 40(15):5916–5923, 2013.

L. Verdoliva. Media forensics and deepfakes: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 14(5):910–932, 2020.

M. Weber, G. Domeniconi, J. Chen, D. K. Weidele, C. Bellei, T. Robinson, and C. E. Leiserson. Anti-money laundering in Bitcoin: Experimenting with graph convolutional networks. In *KDD Workshop on Anomaly Detection in Finance*, 2019.

J. West and M. Bhattacharya. Intelligent financial fraud detection: A comprehensive review. *Computers & Security*, 57:47–66, 2016.

C. Whitrow, D. J. Hand, P. Juszczak, D. Weston, and N. Adams. Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery*, 18(1):30–55, 2009.

S. Xuan, G. Liu, Z. Li, L. Zheng, and S. Wang. Random forest for credit card fraud detection. In *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, pages 1–6, 2018.

Epilogue

The solutions presented in this hybrid academic–industry white paper illustrate how artificial intelligence is beginning to reshape risk management in Swiss private banking. Across documentation, identity verification, and fraud detection, a clear shift is emerging from periodic and rule-based controls toward continuous, data-driven oversight. This evolution is driven by growing regulatory expectations, increasing operational complexity, and the need to manage risk in real time across global client relationships.

For Swiss financial institutions, this transformation is both an opportunity and a responsibility. Advanced analytics and AI can strengthen documentation quality, improve transparency, and enhance the detection of financial crime. At the same time, successful deployment requires robust governance, explainability, and sustained human oversight. Trust, accountability, and regulatory alignment remain fundamental competitive advantages of the Swiss banking model and must be preserved as institutions adopt new technologies.

The Swiss financial centre benefits from close collaboration between universities, banks, regulators, and technology partners. Initiatives such as RiskON, led by the University of Zurich, provide a structured platform for experimentation, knowledge transfer, and talent development, enabling the translation of academic research into practical solutions. This ecosystem approach supports innovation while reinforcing the high standards of stability, credibility, and long-term trust associated with Switzerland.