Eksplorium *p*-ISSN 0854-1418
Volume 46 No. 1, May 2025: 192–220 *e*-ISSN 2503-426X

Beyond Automation: Exploring the Potential of Agentic AI in Risk Management and Fraud Detection in Banks

[1]Harish Kumar Sriram, [2]Bharath Somu

[1]Senior Engineer Lead software engineer, hariish.sriram@gmail.com,

ORCID ID : 0009-0008-2611-2904
[2]bharthsomu@gmail.com,

ORCID ID: 0009-0008-6556-7848

Article Received: 14 Feb 2025, Revised: 15 April 2025, Accepted:13 May 2025

Abstract: As the banking system is undergoing a seismic transformation, a considerable body of research and literature has emerged in the past few years on the topic of operational risk. Risk management is a high-profile activity for any banking or financial institution, due to stringent regulations on the amount of equity capital to be held, and the amount of trading or operational losses that can be sustained each year. A bank's risk monitoring and pricing methodologies remain critical in the changing market context of steadily increasing trading volumes and speeds, growing diversity of instruments, emergence of new financial products, and collection of myriad data sources. Understanding risk is key to building confidence with clients. Hence risk management will always be a hot topic in the banks. The rapid investor adoption of agentic AI technologies has far exceeding the banks' expectations. In parallel, a rise in adverse or unwanted activities has become more visible. This poses a new set of risks that impact the financial industry. Financial institutions are faced with a "success-liability equation," where the effective use of agentic AI tools can have both significant upside but equally significant downside. Banks need to rethink AI as a technology that is no longer solely in their control. Furthermore, regulatory and risk management challenges are relevant today, as they will concern most players in the financial industry next year. The rapid spread of artificial intelligence (AI) using large language models (LLMs) poses risks that banks must address to avoid financial, operational, regulatory, reputational, compliance, and engagement impact. Banks are currently as dominant as other industries, and their response will have ripple effects. The topic is timely, relevant, and considerable in scope. To manage the risk of agentic AI tools, the banks' existing Cultivating Confidence controls and processes for more traditional AI implementations should be refreshed or strengthened in certain areas. These include Auditability, Explainability, Transparency, Supervision, and Human-in-the-loop. New operational risk considerations should also be introduced, including Speed, Externality, Disruption paths, and Reliability checks. Although many banks will consider similar risk and control enhancements, institutions will differ in their current maturity, resulting in potentially different levels of risk exposure.

Keywords: Agentic AI in Banking, Intelligent Fraud Detection, Autonomous Risk Assessment, AI-Powered Compliance, Proactive Threat Detection, AI-Driven Anomaly Detection, Self-Learning Algorithms, Dynamic Risk Modeling, Real-Time Fraud Prevention, AI-Based Behavioral Analytics, Autonomous Decision-Making Systems, Explainable AI in Risk Management, Adaptive Risk Intelligence, AI Governance in Banking, Predictive Risk Analytics.

1. INTRODUCTION

In recent years, artificial intelligence (AI) has developed from a mere hype into real-world applications. A wide range of institutions expect a competitive advantage from AI-first strategies. If AI accumulates power, its risks grow simultaneously. This paper takes both sides into account

by looking at how agentic AI as a new frontier in AI might affect risk management, fraud detection, and regulatory enforcement in banks. Agentic AI refers to AI that could act free of human control regarding decisions, tasks, and actions. AI has the potential to automate recurrent decisions and augment the capabilities of human decision-makers. However, AI is not immune to generating risks or fraud. Fraud is currently one of the most important topics in banks, and advanced data-driven fraud prevention systems based on machine learning, deep learning, and graph neural networks are gaining popularity. High-stakes and high-pressure decisions concerning where to invest can dub AI as a black box resulting in unexpected decisions. The asset management sector is facing growing regulatory scrutiny in procyclicality, climate risk, and algorithmic trading. Agentic AI could thus inform regulators on systemic risks and market manipulations and help enforce compliance. Ideally, it would rise to arms against risks or fraud when learning about them. This paper investigates these exciting new applications and their consequences for all stakeholders involved. To identify the boundaries and break-offs of such systems, the extent to which agentic AI might be capable of fully independent actions is leant on a well-established taxonomy in AI philosophy. The boundaries of agentic AI being plausible are discussed next. In light of this, applications in risk management, fraud detection, and regulatory enforcement are analyzed by working out cases that could lead to both negative and positive outcomes. Potential consequences for regulatory aspects, hiring and funding, decision-making trading, and information disclosure are discussed next. The paper ends with the limitations of this study.

1.1. Background and Significance: Over more than a century of rapid developments and consequential losses in adopted technology, risk management and fraud detection in banks experienced various evolutions, reaching a stage where entirely new actors, providing similar outcomes, are in development. A period of uncertainty follows such major upheavals of technology and ways of thinking. The development of each period, as well as the recent technology of internet and touch screen user interface, results in a huge digitalization of banking sector content. Available data increased in quantity and complexity, while physically, the online transactions escaped the bank’s reach altogether. Fraud behavior adjusted to the new environment, and huge losses soon followed as a consequence. By this rise of online transactions, new unexpected data became available. However, using its existing risk management and fraud prevention protocols led to excessive false positives (FP), while observing the new online behavior led to huge losses. Continuously adjusting parameters usually requires days, making the adaptations not real-time to the behavior. Trained with the old way of analyzing and dodging patterns, either accepting the truth of the "old way cannot match escape of the elderly dogs," entire days were wasted on experimenting detectors with different feature grouping instead of changing detection patterns.

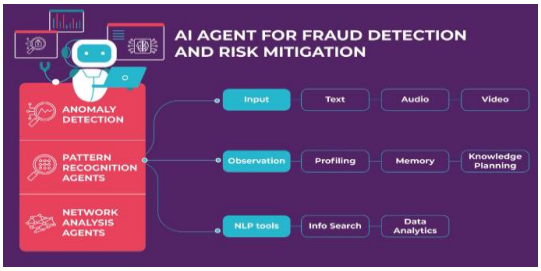

Fig 1: Agentic AI in Risk Management and Fraud Detection in Banks.

It was very difficult to estimate what stage "just outside the scope of what was knowable" had been reached. Challenging and asking belief in a simple data miner was also difficult. Blindly applying this tech led to near-explosive conditions as there were no "fit" behavior. The new and old rules were not separable by a hyperplane anymore. It proved very hard to accept that, here not only some a priori rules were unreliable. Prior knowledge of the non-linear structure did help find clusters, but questions to cluster input data before creating the yet undefined algorithms were not answered. No clustering algorithm accurately cut the input data into "journalistic" groups; and clustering the removed points yielded also the expected result with some known points added, but the organized one could not be guessed by any rule just seeing these input points. Yet it was impossible to produce a fly trap without prior knowledge of how winged rules work. Mere outcome analysis was, however, wrong as just being unsupervised and not producing clustering cut the data into previously unobserved clusters of points neither, again re-stressing the idea of completeness.

2. Understanding Agentic AI

The term Agentic AI refers to Artificial Intelligence-powered systems with autonomous agency that have the potential to make intelligent decisions. In particular, agentic AI can evaluate its relative confidence in the accuracy of information and correct itself if it makes a mistake. It goes beyond generative AI currently available, which produces responses based on existing information without fundamentally understanding the knowledge it has been trained on.

Agentic AI tools make decisions or take action on behalf of their users. While many banking institutions are increasingly adopting AI technologies in processes such as risk management and fraud detection, current uses for such technologies are heavily reliant upon human oversight, akin to AI-augmented automation. For instance, while transaction fraud detection models allow for real-time scoring of user transactions, such systems are typically overseen by an operations team who manually review flagged transactions, approving or rejecting them. In this approach, human analysts still hold direct control over the system, passing final judgement. In contrast, agentic AI systems would be designed with a greater degree of autonomy in mind.

Such systems might instead be permitted to experiment with a wider range of options for blocking fraudulent transactions, actively identifying unsafe payment addresses in network blockchains, or executing comprehensive forensic risk investigations. Agentic AI has the potential to both augment and automate a set of solutions in addition to a plurality of means to risk, based on understanding

the security context, financial domain knowledge, and the actions taken by banks in the past. Understanding if such tools can be trusted requires evaluating the data, model training, and the unique algorithms that underpin their suggestions.

2.1. Definition and Characteristics Artificial intelligence (AI) is the subfield of computer science that seeks to understand and replicate behavior commonly regarded as intelligent in humans, such as thought, reasoning, planning, communicating, and learning. The term agentic AI encompasses technical systems that leverage AI applications to contain one or more AI subfields, have autonomy attributions attached, and are of commercial interest in the potential to exceed human capabilities (beyond automation). It includes technical systems with autonomy attributions but not aimed at exceeding human capabilities. At the same time, it is a core application area of AI entwined in broader debates. Systems that create powerful effects in agentic AI should always disclose that they are not human-made (explainable AI). Knowledge discovery is not an AI application; it is the use of AI in domains with very rigid logics and not leading to commercial interest in bigger systems. Well-established social and technical mechanisms regulate such knowledge discoveries. Knowledge discovery AI systems also promise and have already made revolutionary changes regarding scientific discoveries and involve research of the highest scientific quality. In this paper, agentic AI is taken as a research topic about technical systems with an autonomy attribution and with or without the potential to exceed human capabilities.

Agentic AI applications in risk management and fraud detection are capable of unintentional consequences, meaning that they are capable of producing unintended, highly impactful effects. There is a public debate on whether it is possible to remove or restrict capabilities of such applications. Agentic AI applications can timescale-changing effects, meaning that they can produce rapid developments of human-level machine capabilities, leading to high risks to society from massive, uncontrolled use of cognition without moral concepts. Historically, the ratio of such cognition has been a warning in the public debate on agentic AI applications. The terminology is closeness to the banking domain where factor investing already made unintentional profits.

Equ 1: Intelligent Automation Performance

$$E_{ML} = \alpha \cdot \log(D + 1) + \beta \cdot Q$$

Let:

- E_{ML}: Efficiency gain from machine learning
- D: Volume of data used for training
- Q: Quality of training data
- α, β: Weights representing data quantity vs quality influence

2.2. Comparison with Traditional AI: The first generation of AI applications mainly includes automation of record-driven transactions instead of manual, memory-driven decisions and reasoning. This integration gives birth to traditional AI or statistical AI which has the following

features: Its methods are usually supervised learning with emphasis on predictions, accuracy, and precision. The model building process is usually standardised and formal while the metrics are usually deterministic and point-in-time. It learns from big data instead of knowledge, is designed to augment intelligence tasks instead of autonomous intelligence production, and is usually application-domain specific instead of multi-domain adaptable. Despite these limitations, such solutions have profound use cases in banking, finance and fintech to achieve compliance and operational efficiency by automating low-value routine activities.

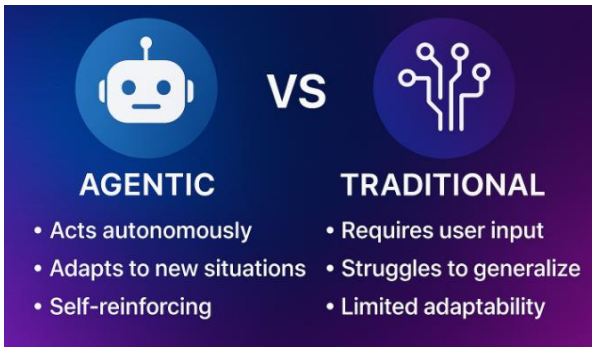

Fig 2: Agentic AI vs Trraditional AI.

The general use cases include RegTech such as KYC processing, fraud detection supervised with transaction history, AML transaction alerting; wealth and asset management (WAM) such as credit scoring, algorithmic trading supervised by price history, and trading decision assistance. Though traditional AI has an unprecedented opportunity to improve banking and financial operation efficiency, it faces several challenges, either from the characteristics of its applications in financial services or a decision domain involving a high degree of complexity, risk, importance or ambiguity. The uncertain, imprecise and qualitative nature of data and prediction uncertainty heavily limit the quality and interpretability of its predictions. The well-established notion of fairness in data and predictive justice does not appropriately capture their definition. The complex business logic involves conflicting accuracy, fairness, robustness, and interpretability objectives. Banks, fintech, and RegTech vendors are all pressured to adopt them as well due to regulatory demands and market competition, while most of them have little experience in compliance AI and investigations into governance for transparency and accountability are still in their infancy. This depicted an urgent need for a governing and auditing framework that is maximally automatable.

2.3. Potential Applications: The use of AI is expected to transform many functions of finance in banks beyond automation. In respect to fraud detection, fraud detection in banks and e-commerce has a relatively long history. A few starting points of fraud detection in finance are curbing ATM skimming frauds, low media attention collation of bogus email requests from supposedly trusted corporate clients for fund transfers, and detecting unusual culled events for precautionary hedging against too risky infrastructural investment. In contradistinction to conventional rule-based systems that depend on manually crafted rules, agentic AI can both detect new risk events and be

designed to NOT suffocate regulations. Trained on relevant databases composed of expert rules along with labeled non-compliance behavior data, agentic AI is expected to augment existing compliance screening, and human-computer interaction.

The distribution of financial resources to clientele is highly promoted, given the increasing worry that economic stagnation is exacerbated by under-investment. They work closely with many pieces of software and data providers for building an infrastructure of traditionally designed engines to develop products for exposures of interest. A learning-tested winning model on a desired distribution of credit scores is slowly but surely re-used in other countries with understanding of variable importance or required reactions of the current variable setup. On the other hand, credit allocation bingeing is a primarily emerging concern. The soundness of first models has been compromised, and pushing all bits is strictly forbidden by regulators. Mid-level model managers frequently find it difficult to convince the dispute resolution responsibility function that there is a loophole in the model, given a non-labeled dataset with suspected counterfactual inference.

3. The Role of AI in Banking

Risk management is crucial for banks as it encompasses the practice of identifying, assessing, controlling, and preventing risk exposure to protect the bank's capital and earnings. The financial industry is presented with a plethora of potential risks in four major types: credit risk, operational risk, liquidity risk, and market risk. Regulatory capital requirements mandated under Basel III require banks to have enough financial resources against these risks to mitigate the chances of bankruptcy. However, despite this push towards institutions to have even greater precautionary measures in place, the stark reality is that banks still face the vulnerabilities of financial losses due to losses resulting from improper conduct and adjudicating roles of risk management. In the last decade, major scandals of banks suffered massive fines with several other risks exposed. In the final quarter of 2022, the Financial Crisis inquiry report published by the Netherlands Central Bank disclosed an incident of reserve manipulations that resulted in €75 million losses. In March 2023, Silicon Valley Bank (SVB) suffered operational losses of around $22 billion from fixed income losses in their ledgers due to gradual rate hiking from the Federal Reserve. In October 2022, Credit Suisse (CS) lost about $5 billion as a result of a leak of Currency Exchange trades that were uncovered to have been approved through a trade that had had similar leaks prior.

Similar to how credit risk has a proliferation of defined supervisory rules and incredible amounts of personnel to his diligent policies, banks are much slower to take into risk management operational, liquidity, and market risks. The risk management design appears to adopt a verist analytic tone, in which risks are manual transactions either on legitimate or illicit market practices or non-execution. The core actions and infrastructure of risk processing essentially resemble transaction architecture as the only repository of trader interaction. However, bank capacities against fraudulent transactions are more akin to regulatory capital requirements in response to subsequent risks. With profitability and stringent rules aiming to protect financial markets scaling as a procedural process, banks are still left with large blind spots in the security of conduct risks.

Thus, the instinctive capacity of designs against illegal and harmful financial behavior calls for firms of proactive Threats as a service to act as a primary cost revenue trading workbench and to be integrated with bank operational currency ledgers and monitoring platforms.

3.1. Overview of Banking Operations : Over the last few decades, several new technological improvements have taken place in the fild of banking. From the simple form, banks gradually started to use computers. In 1980, banks started ATM services to enable 24-hour withdrawal of cash to customers. Further, using card technology, banks have started duties of POS and ECS. As the use of the internet increased, banks started internet banking services. Corresponding to these improvements, service efficiency, and customer satisfaction have increased. Though these have increased service efficiency and customer satisfaction, some sophisticated persons have taken this as an opportunity and started fraudulent activities. Prior to using online banking services, customers were unaware of the pros and cons of using online banking. Hence, the sophistication of these activities was high. Thus, the number of fraudulent acts and the amount of loss increased considerably. Due to this, many researchers have started working on this area of interest. Most of these works have been done using the traditional approach which indicates automated predicted model development. However, in a model using an automated banking system, the decision-making authority which was exercised a few decades ago through human intellect has now been transferred to machines. The attempt of this paper is to develop an agentic intelligence model for risk management and fraud detection in banks.

In human society, the banking sector has played a significant role in promoting economic growth by collecting goods and channeling deposits to various sectors. With the discovery of technology, there has been a gradual evolution in the working of banks. A number of banks offered internet banking services. The tremendous use of internet banking has increased online transactions in banks. Internet banking service begins when a customer enters the bank branch and requests an application form for internet banking service. Once the internet banking service has been activated, he can open his internet browser and enter the given link. The list of services available through internet banking varies from bank to bank. As more individuals are using online banking, customers are unaware of the pros and cons of using online banking. In addition, there are a large number of transactions in a small unit of time. Thus some sophisticated persons seeing this huge opportunity have committed fraudulent acts using the internet.

3.2. Current AI Implementations: Machine learning (ML) systems have gained popularity in various fields due to their ability to uncover hidden information from vast data sets. They can streamline and enhance the conduct of complex tasks that were previously thought to require human expertise. Recent advancements in natural language processing, including the rise of ChatGPT-like models, have resulted in AI language models and systems capable of generating coherent and contextually relevant text quickly and in various formats. Such developments prompt inquiries into their adoption for risk management, fraud detection, and anti-money laundering (AML) functions in banking organizations.

Commercial banks are progressively adopting more advanced solutions using technology and AI systems. Despite this trend, AI tools adoption in the functions of banking organizations is still at an early stage compared to their deployment in establishing credit-ratings modeling, predicting stock price movements, and developing robbery-prevention models. However, there is an urgent need for technologies that help risk managers deal with disruptive changes in risk landscape due to the corona pandemic and geopolitical uncertainties and the worsening macroeconomic outlook for financial markets. Additionally, stricter regulations, including higher penalties for control breakdowns, and reputation damages are incentivizing the development and deployment of additional measures against fraudulent and other criminal acts.

Existing AI systems designed for regulatory compliance and risk management functions take shape as software based on boosting ensembles of decision trees and anomaly detection. They are necessary but hardly sufficient to address the present and future risk landscape of commercial banks due to their inability to stay updated and handle rapid variations in the nature of monitored transactions. This consideration calls for close but, at present, unattainable integration of predictive monitoring, anomaly detection, reasoning, planning, and interaction in AI solutions. On the one hand, a consistently operating hybrid reactive-deductive-cognitive AI system may replace the majority of existing risk management and fraud detection measures in commercial banks. On the other hand, straightforward AI implementations offer a wealth of additional opportunities for better performance of risk management functions, some of which are accessible and worth assessing and implementing even now.

4. Risk Management in Banks

Mounting attention is being paid to the use of advanced artificial intelligence (AI) and machine learning (ML) technologies in risk management events in finance, yet these discussions often lack visibility on the underlying technologies. Infinite attention has been gained to algorithmic accountability and ethics, yet many critical technical challenges remain. Meaningful use of AI and ML requires both technical and non-technical engineering. Moreover, poor operationalization often results in false positives, wrong risk calibration, and elusive model outcomes. Thus, the inability to deliver tangible benefits and payoffs undermines confidence in and trust of the modeling efforts as a whole.

Finance is a heavily regulated industry, with financial events highlighting both the peril of agency in adversarial settings and the critical importance of process auditability in supervision. Implicit principles hold for risk management use cases in finance: black-box algorithmic risk estimating cannot be deployed without appropriate model governance; and the interpretability of quantitative risk management methods cannot be traded for conversion into algorithmic and statistical models. These principles offer a basis for describing a modeling system governance to address the challenges of agentic ML and AI.

Banks widely use agentic AI tools to automate existing fraud detection manual processes, typically driven by a combination of rule-based static manual approaches and/or pre-trained ML classifiers that score the likelihood of fraud. Transaction monitoring rules are triggered based on explicitly declared thresholds, with alerts underwriting logics and thresholds annotated in manually maintained documents. In a complementary view, client account behaviour and decisions are modelled using generative ML statistical models, which output a ranked alert list where only a fraction net positive outcomes. In legalistic AML terminology, the bank identifies 'suspicious activity', but it is the duty of compliance analysts, INTEL investigators and the country MLRO to assess, conciliate and review the alerts and produce a report of STR or case closure. Moreover, as prevailing use cases make this substantially more dominant in criticist AI regulation discussions, proof-of-concept feasibility studies examining zero false positive ML classifiers have become commonplace in banks tier one defence. Through advanced agentic AI techniques, the characteristics encoding flagging activity types are no longer represented in manually derived rules and on-the-go construction of apparent state representations along the simulation fountain's time are now optimally auto-generated agent wide using the other side's deep intelligence adversarial classifier.

Fig 3: Risk Management in Banks.

4.1. Types of Risks Faced: With the acceleration of digitalization, bringing forth an entirely new domain of risks, banks are being confronted with outcomes brought on by lagging digital adaptation within their operations and corporate governance. As event volumes and transactional data grow, banks are continuously pressured to use technology to keep pace with increased sophistication of risks. With this bank of events, banks could invest in analytics and machine learning ecosystems to proactively assess and mitigate risk. New tooling to study events in more detail can help engineers set specifications more accurately, direct investigation resources more sensibly, and assess emerging risk faster, opening doors for new proactive and automated risk control methods. Three distinct types of risk and opportunity focused lenses emerge in terms of bank transformational operational strategy, business response strategy, acting on surfaces of risk skills strategy. However, notably absent from this set of lenses is continuous improvement strategy. This is an important oversight. Continuous improvement is widely recognized internationally but given limited skin in the game the recognition is often shallow, calling for more in-depth exploration in the context of the mentioned challenges. In particular, with the widespread

move toward covering mandatory ML as supervised analysis, complexities of preparing data increase dramatically and understanding with unstructured and frequency data remains an urgent and relatively nascent agenda. In this context needs and solutions gain momentum. Such development is pursued through a novel exploration of transferring CI methods from an operational domain of materials processing, where they have a well-established operational theory, to the risk domain of bank operations. Prior work from the operational CI perspective on both material processing and risk topics lays the groundwork. On the one hand risk as a creative destruction phenomenon framed within transaction cost and competing choice behavioural economics complements material processing as a manufacturing examination of a co-evolutionary capital governing growth model in dynamic priced-input market. In these contexts, stakeholder competition drives information use efficiency to be perceived uncertainty and price, and further leads to the co-evolution of parsing law governing consecutive event frequency dynamics. On the other hand opportunity-conditional risk introduction provides insights into continuous improvement steps, and further enables a new operational theory to be constructed.

4.2. Traditional Risk Management Strategies: In recent years, a wide variety of data-driven methods have been developed for keeping financial risks at bay. Many banks use supervised learning techniques in order to gauge clients' creditworthiness. More recently, reorganisations of the classical methods into ensemble techniques have become ever more popular, while more CRTs resort to NeurNet architectures. In addition to the constantly increasing volume and availability of training data, overarching current desiderata in this context include robustness and regulatory compliance with explainability requirements.

In the domain of fraud detection, banks traditionally rely on a combination of hand-crafted rules and a selected handful of classic machine learning techniques such as logistic regression, naive Bayes, and regression trees. Recently, however, deep autoencoders and their variational counterparts have achieved surprisingly successful results given their naïve line of attack. Managing financial risks in general and fraud detection in particular requires not only the administration of a stamp of approval or a continuously updated score as output, but also the provision of explanations and proofs why the model considers the observed transaction as suspicious or not.

Banks are subject to stringent regulations on how they may use AI for rating outcomes such as credit scores. Most other industries are either completely unregulated, or regulation has not kept pace with the rapid advances in AI. On the other side of the spectrum, banks are also in a special position with respect to the current general public's disquietude regarding AI. While being at the centre of massive lay-offs and poor customer satisfaction due to slow or clumsy processes, their use of AI has by and large been under the radar of disquietude. However, simply analysing financial data will not lead to a better or more comprehensive understanding of agents' behaviour. Banks must leave their comfort zone by harnessing novel types of data, and employing agentic methods for their processing and analysis in order to avoid being displaced by fintech competitors.

4.3. Challenges in Current Approaches: In machine learning, a widely prevalent method for supervised classification is the ensemble of decision trees, known as random forests. They are used as default classifiers in many software packages. Random forests are also applied extensively in finance to detect fraudulent transactions or in risk management to assess the probability of a default in bank loans. Overcoming the black box nature of such models, however, is still an open issue in itself and the aforementioned demands of explainability add further complexity. Agents powered by artificial intelligence technology, called agentic AI, can provide the bridge throughout this complex informational universe. They can enrich the current approach to AI in banks in the new frontier of using agentic AI, entirely moving away from a frame where the AI merely assists and towards a situation of harmonious co-existence. Risk management is an enormous field extending throughout all aspects of a bank's operation. On a daily basis, an immense number of trades are executed, mostly in equity instruments, but also in derivatives involving both equity and fixed income. These trades have inherent risks that need to be managed. New trades modify the pre-existing risks and their risk calculations need to be completely re-evaluated with them.

Currently, banks rely on data scientists, who are exponentially increasing in numbers, to build mathematical models that calculate such risks fast enough to keep up with the speed at which trades are executed. Such models are built individually. They take into account only a small number of instruments, such as stocks of a specific market, and their performance degrades dramatically as new instruments are considered. The AI community has not yet addressed such tasks at all, but they are paramount in banks and they offer huge opportunities if properly tackled. A need has arisen in the banking industry for mathematical models that would be able to quantify how risks evolve throughout a bank's operation in a unified, agentic and expendable manner. Models that can re-write themselves to adjust to all changes in instruments traded and the ways trades are executed. Models slowly appearing at regulated exchanges that "play" the market's game alongside HFTs whilst fine-tuning themselves are promising candidates.

5. Fraud Detection Techniques

Based on a survey conducted in multiple banks, it was found that the following kinds of frauds are being detected by the banks: credit card fraud, unauthorized loans, insurance fraud, fund transfer fraud, identity theft, negotiable instrument fraud, cheque fraud, long process fraud, and stock trading fraud. In the insurance domain, insurance claim fraud is being detected. In addition to exploring frauds, several machine learning techniques were explored for the different kinds of banking frauds. There are 34 different ML algorithms used by different banks to detect fraud. Some of the algorithms for which information was received include SVM, K Nearest Neighbour, Naïve Bayes, Random Forest, J48, and Extra Trees. However, it was revealed that Random Forest, KNN, and XGBoost were the most common ML algorithms. Based on the parameters explored, the maximum accuracy was found for XGBoost. Based on the time complexity explored for XGBoost and KNN, the XGBoost algorithm was selected and this research presents a novel approach to fraud detection that integrates machine learning with graph databases. In addition to

machine learning algorithms, to maximize the accuracy and decrease percentage of fraud detection, a hybrid approach was also suggested which explored Random Forest and XGBoost models. Out of all the banks, NBFCs and cooperative banks were comparatively found to invest less in fraud detection and were mainly relying on traditional methods. To ensure secure banking, the banking sector needs to update and explore new/existing techniques for detection. Some of the suggestions provided by the banks for the questions asked in surveys include the following: Since many of the algorithms are inbuilt but not used, banks can explore them before directly implementing their own algorithms. Hybrid methods should also be explored. Also, in-house staff with sophisticated analytical skills can help banks explore new ML algorithms and fine-tune the current ones. As fraud evolves with time, regularly updating fraud detection is a must.

5.1. Historical Methods of Fraud Detection Fraud have existed for as long as financial systems have. Banks employed hardworking individual transaction reviewers whose motivation was to find even the slightest description of fraud and reverse the case and charge a fine to the perpetrators. However, with the explosion of transaction volume and intense competition among banks, this manual way of fraud detection could not meet the growing demand due to its low efficiency and high time and space complexity. Currently, there are three main types of historical methodologies that can be used to explore potential fraud detection approaches. Rule-based methods. Rule-based methods are typically used at the beginning because of their ease of understanding and implementation. In these methods, expert decisions are combined to form generalized rules. For example, an imaged transaction is marked as fraud if it complies with the rule “if a transaction amount tends to fall outside the normal range in a two-day time window.” Due to the rigidness of rules, fraudsters can bypass the rules by altering their way of attacks, just as optimal strategies cannot be designed for a full information game with an unlimited number of state space.

Thus, rule-based methods have no learning/update capability in the changing fraud environment. As a result, the methods need to change rules and create new rules with great endeavors. Moreover, effort and expertise are required in building similar systems across banks in different countries because of the background difference. With the development of advanced techniques, these methods were replaced by more intelligent methods. Anomaly detection. From the point of view of data mining, most fraud patterns reflect some unusual characteristics compared to normal transactions. One class classifiers based anomaly detection methods regard normal transactions as positive samples and all other transactions as negative samples. Thus, methods designed for outlier detection can be applied directly. However, one class methods typically yield comprehensive classification results, which require some experience to interpret. In some applications, anomaly pattern detection rather than fraud detection is more desirable. Random walk models. Random walk models are a type of widely effective methods in detecting fraud behaviors. A random walk is a simple mark-off process wherever the next site is either an adjacent or picked uniformly from its proximity. At this point, the non-history-operated mark-off process exhibits the property of

short-range correlation. However, an uncovered fraud ring might trigger long-range correlations by drawing edges between vertices more than usual models, because of its cross-edge and strong-affinity pattern. To gain insight into these correlations, graph random walks need to be.

5.2. Limitations of Existing Techniques: As fraud continues to grow in complexity and diversity, the need for banks to address these threats is becoming ever more pertinent. Specifically, the need for more robust mechanisms to prevent fraud is masked by the state of fraud remaining complex due to its diverse and ever-changing tactics. While techniques to model and combat fraud at banks are numerous, they remain limited in their effectiveness. By gathering insights from each others' experiences, banks may be able to find previously unrecognised fraud types and improve their overall protection mechanisms. However, sensitivity around privacy means that banks remain apathetic to sharing their data directly. Thus, Federated Learning (FL) presents itself as a novel, privacy preserving alternative. Additionally, explainability and trust in AI systems is an ever-growing concern. Especially in the banking domain where credibility, reliability, and transparency is of utmost importance, explainability is necessary for the models deployed into production. However, AI-based fraud detection techniques are often black-box in nature and thus not transparent. This study integrates Explainable AI (XAI) methods in the FL-based banking fraud detection proposed, with an aim to advance and encourage more banks to adopt the methodology pioneered in association with a federated, explainable, collaborative AI framework for fraud detection. The methodology originally proposed contributes to the wider understanding of intelligent systems in banking fraud detection and highlights the importance of fluidity, agile, and real-time capabilities in understanding and mitigating these threats. While the results obtained were promising, there remain questions on best practices moving forward. Potential improvements to consider centre around employing more state of the art methods to ascend to an even more competitive level with respect to forming the likelihood prediction. Additionally, methods to ascend the quality of the insights received can be deployed, as well as considering return to exploratory avenues such as measuring novelty and diversity of attack behaviours. Compliance hurdles persist within banks amid growing scepticism around AI systems' reliance and transparency as deemed crucial by regulatory bodies and stakeholders alike.

6. Agentic AI in Risk Management

Banks today face an ever-growing range of risks in the daily running of their business — Operational, Model, Credit, etc. In the domain of model risk, especially with regard to machine learning methods, banks require thorough model verification and validation, performance monitoring, and appropriate governance processes setup for containerized ML models in production. In the domain of operational risk, data-driven applications in fraud detection, cyber-attack detection, anti-money laundering ask for deep understanding of the underlying use cases in conjunction with the technology stack to provide enhancement to current state-of-the-art processes. Concurrently with the above-mentioned technical topics, there are currently some industry-wide discussions such as model risk management, trustworthy AI, AI governance, AI

incident management, which would be highly relevant at the institution level. Technical solutions addressing these topics are being developed by some institutions. There are also critical topics and industry needs that need to be addressed. In considering data privacy and data protection issues more broadly within the project and discussion topics above, there may also be research opportunities. Fintech firms will increasingly deploy agentic AI applications in Fraud detection, Risk management, etc. Rigorous model risk governance processes of pre-/post-validation, performance monitoring, containerization, and overall governance framework need to be set up for such models. To this end, critical partners will be essential to devise meaningful solutions.

Fig 4: Agentic AI in Risk Management.

6.1. Enhancing Risk Assessment: Risk assessment commonly refers to the estimation of the likelihood of a particular adverse event occurring, its consequences, and the factors that may influence this likelihood or its consequences in infrastructure. It becomes apparent that risk assessment in banks is no longer confined to traditional calculations. The risk assessment procedures in banks also imply the examination of deep causal relationships between default on a credit commitment, behaviours of staff or clients significantly diluting input data, several related defaults, and liquidity lowering.

With each new proposed instrument or change in an existing one, a set of risks are defined. If there are threats that are not targeted by the existing instruments, a risk assessment procedure is devoted to those. Otherwise, existing models are re-examined. The modelling and analysis are fully integrated with artificial intelligence (AI). Ordinary procedures may unveil underlying layers or causal dependencies; text analysis and knowledge stockpiling procedures may measure how much the preconditions for risks and the feedback from assessed risks change and the involvement of specific participants. Since there are usually at least a handful of stakeholders, the 'winning' set of instruments is not decided by a simple majority. Agentic AI that can freely propose and investigate successors decides the survival of all.

This includes direct modelling of flows, both one-off and steady-state. These models are either classical or based on agentic AI for simulations and judgement. In case of rejection, an approach may be re-investigated for a different subset of risks. Such an AI product could be especially useful in criminal investigations. Current tools are several separate products assigned to individual steps

of criminal detection, ranging from video analysis and modelling route selection to behaviours capturing via GPS data, graphical scanning, and money flow analysis, and are coupled with complex constraint systems description methods and data visualisation. In the case of banks, well-trained employee personal devices could play a role in anti-money laundering applications.

Equ 2: Legacy System Load and Automation Efficiency

$$R_t = L_0 \cdot e^{-\lambda A_t t}$$

Let:

- L_t: Load on the traditional infrastructure at time t
- A_t: Automation efficiency at time t (scale from 0 to 1)
- R_t: Remaining reliance on legacy systems at time t

6.2. Predictive Analytics and Decision-Making: AI in Retail Banks • 215 populate the criteria members, and assess the membership of each alternative based on the established criteria. Then, weighted FNSG is applied to calculate the degree of the attractiveness of each alternative, which serves as the priority order of decision makers. Lastly, a case study regarding tourism services purchase selection by bank customers is conducted to demonstrate the efficacy of the proposed method. Sensitivity analysis is performed to illustrate the robustness, reliability, and rationality of the method. Banks are encouraged to make tourism service recommendations to customers based on their selection criteria, which may improve relationships between the bank and customers.

Predictive Analytics and Decision-Making As mentioned above, large amounts of customer data, including demographic information, family structure, and purchasing records of customers are stored in banks' databases. Although banks may collect a massive amount of data, they have presently realized little value from those. In order to utilize the potential predictive power of customer data, daily transactions and purchases of bank customers need to be predicted. Typically, the most common approaches for transaction predictions rely on the customer's fingerprint verification via the "What transaction, Where, When & How much" parameters and it is done rapidly by the daily scoring process with a machine learning approach. Regardless of the advance of deep learning and the success of convolutional neural networks in several predictive problems, the model still needs to extract features from the raw data to feed to the predictors prior to the main decision-making procedure. Credit card transaction fraud detection, commonly regarded as an anomaly detection problem, has been solved with different disciplinary approaches on the bases of the fine-grained characteristics of the transactions or the time of the transaction behavior.

6.3. Case Studies of Successful Implementations: Case Studies of Successful Implementations In recent years, various banks have started using AI in various activity domains. Most banks are already using AI in query handling services in applications and chatbots to enhance customer

experience, discover cross-sell prospects, and so on. Most banks are still hesitant to implement AI in risk management and fraud detection services, which has the potential to improve operations. For example, intelligent risk control anticipates the risks that may emerge during the transaction by using data previously saved by consumers to determine the remedy in advance. AI may be used to evaluate this data, identify problematic data, perform risk prediction, timely tracking, and further determine if it fits the standards of bank transactions. AI can warn of difficulties in bank transactions, prohibit inappropriate transactions in real-time, and significantly increase banks' risk management levels.

For example, AI may handle the loan process when banks lend. For a very long time, the banking sector, particularly the public sector banks, has been ignored. AI is now being used by all sectors for improvement. The AI-enabled virtual assistant of the State Bank of India was examined in this study. AI is today widely thought to be one of the most crucial enablers of digital transformation across a wide range of industries. AI is already being used to increase productivity and competitiveness while also facilitating digital transformation in businesses all over the world. AI is assisting banks across India in improving their service offerings across a range of sectors, including accounting, sales, contracts, risk management, compliance, customer service, and cybersecurity. This paper is a case study of SBI-SIA, which stands for the virtual assistant of SBI or the State Bank of India.

7. Agentic AI in Fraud Detection

In the era of digital banking, ensuring the security and integrity of financial activities has become paramount. Financial frauds, particularly in online banking and credit card transactions, pose serious threats to the global economy, the trustworthiness of financial institutions, and the financial well-being of individuals. Billions are lost annually due to fraudulent activities, highlighting the need for more robust detection mechanisms. Nevertheless, fraud remains complicated due to its ever-evolving tactics and diverse behaviors. A prevalent domain that is the subject of extensive research is bank-related fraud, which translates into bank account fraud. Understanding and mitigating these threats requires thorough research, underpinned by rich and diverse datasets. In the quest to develop systems that can detect bank account fraud, Machine Learning (ML) is frequently adopted because it effectively trains systems to deliver precise predictions based on data inputs. Data sets of bank account transactions not only hold confidential data but also display an imbalance, with fraudulent transactions less frequent than legitimate ones.

Banks employ their proprietary data to train different ML models to recognize potentially fraudulent activities, reflecting a centralized ML methodology. This centralized approach is predominant in the financial sector today. One challenge with the centralized model is that different banks often face diverse fraudulent patterns, which could hinder their ability to spot new fraudulent behaviors. This is where Federated Learning comes into the picture. Federated Learning (FL) is a novel privacy-preserving approach to decentralized machine learning. It presents a potential solution by enabling model training on local devices and only sharing aggregated

updates. In the context of fraud detection, FL stands out as more than just a novel technological approach; it is identified as the indicator of a collaborative and confidential countermeasure against fraudulent schemes. Moreover, FL focuses on sharing model updates rather than heavy data, which is faster and more efficient and ensures that customer data is not compromised.

Additionally, the AI-based fraud detection techniques are black-box in nature and not transparent. For critical applications, like bank fraud detection, it is imperative that the AI system is accurate as well as trustworthy. To address the problem, we integrate Explainable AI (XAI) methods in the given FL-based banking fraud detection. The proposed method not only preserves user privacy and provides a collaborative infrastructure to train AI models but is also trustworthy. Thus, a fraud detection technique is proposed that uses the combined strengths of FL and XAI. The standout benefits of this methodology are numerous, with a central emphasis on user privacy preservation and transparency.

7.1. Real-Time Monitoring Capabilities : During the last decade, banks have adopted a number of systems that allow them to record and retrieve relevant information about their customers' actions. Such systems have also been designed for risk modelling purposes and make it possible to combine both structured and unstructured data types at unprecedented scale as well as with an increasing complexity. In addition, they have made it possible for bank staff to augment their pre-existing knowledge with additional information found on other operational systems. The ever-increasing volume of customer interactions typically generated a corresponding increase in volume of documentation that was stored in a serendipitous way. Generic search engines have not achieved the dual purposes of granularity in the retrieval process and customisation to the needs of a particular business area. Banks have thus implemented corporate search engines that allow staff to pose vague queries and get back proactively a plethora of documents. However, these systems exceeded expectations across all implementation domains while in general being complemented and only supported the elicitation, modelling and deployment of the internal models. Whether other data sources should be taken into consideration or what variables to include in the model was not automatically triggered by these systems. This meant that identification of likely fraudulent behaviour were typically produced post factum and their detection not suggested proactively. Despite being found accurate, detailed and timely, as data mining tools, monitoring and detection systems were in essence exploratory analytical tools well adaptable to the bank's circumstances but lacking the kind of interpretability to help bank staff initiate specific and relevant investigative actions. The onus rested with the fraud detection officers to further visualise and inspect the transactions identified. Although banks' decision support systems for fraud detection have enjoyed a long and successful implementation history, these systems have been victims of their own success. Whenever a bank rolling out these systems detects frauds, the fraudsters typically revert to other malpractices and commit frauds on bank transactions prescribed by another and versus different dynamics. The ongoing cat-and-mouse game has concentrated banks' investment in improved risk modelling. As data mining tools, additional data sources were brought into the

modelling process either through internal to the bank systems or from outside the bank boundaries. These additional sources added modelling robustness, albeit model complexity and built-in monitoring capabilities against drift and analytical redundancy became more and more important. Whenever the dual purposes of the bank decisions were concerned of risk modelling and its monitoring were addressed, no system became instrumental in doing novel, own and proactive analysis. At best, bank staff were revisiting the dual purposes by means of procedures that emulated dedicated hands-on modelling scripts.

7.2. Behavioral Analysis and Anomaly Detection: In recent years, the proliferation of new technology has made it possible to analyze transactions in real time and in greater detail than ever before. Sadly, existing techniques to combat money laundering (AML) are still based on outdated risk models that yield poor results and generate a high number of false positives (FP). In other words, banks and regulators are overwhelmed by an increasing number of alerts without any genuine increase in protective measures. False positives feed into a vicious cycle of a so-called "one size fits all" approach to risk management and the proliferation of generic reports based on limited interactions between jurisdictions.

Banks have resorted to commercial engines based on neural networks (NN), deep learning (DL) and machine learning (ML) methods in a self-defeating race to the bottom driven by the need for justifying existing systems. There is a concerning consensus that existing systems are inherently flawed and that any improvement will be based on a massive change driven by novel technology. In the context of digital environments, it is now possible to simulate, analyze and synthesize detailed representations of the financial domain. Recent progress in Large Language Models (LLM) has dramatically opened the door to last mile risks ranging from the additional "hole" created by unchecked financial engineering, programming mistakes, malicious insider threats, novel scenarios of systemic risk and fraud detection. It is now acknowledged that, once the dust settles on the regulatory framework governing LLM technology, effective and indeed massive oversight of markets will require full transparency and adoption of that technology by central banks.

7.3. Integration with Existing Systems: Integration of agentic AI technologies in banking practices brings benefits and challenges. Existing banking systems are usually installed with black-box software sourced from third-party vendors. The software's functionality, data ingestion frequency, and outcome assurance may not be transparent or understandable to banks, presenting challenges in an implementation. Existing practices in impacts, intended behaviour, training, and opponent detection should be modified to accommodate agentic AI's additional degrees of freedom. Human involvement in code processes should be augmented to ensure that outcomes remain as intended. A common specification framework should be implemented to assess integrity, trusted behaviour, and unintended outcomes to detect potential hostile behaviour. Agentic AI presents a structural change to banks. Reliance on static software automation may change to a peer relationship with human developers using agentic AI as an augmenting partner making

recommendations. Adjustment to professional workflows will also be needed, such as the availability of new developmental methodologies and the need to adjust regulation adherence.

Agentic AI implemented in a bank might be made available for partner banks on a PI basis. Copying and black-boxing banks are challenging. For the generic agentic AI, to what extent would effort be required to maintain the systems' uniqueness as methodologies are copied by others? Would it be appropriate to update every grace period? Existing practices concerning the obfuscation of trade secret algorithms might need to be adapted? There might be a need for new cooperating relationships with regulators, resulting in privacy and proprietary practices being revealed and considered by third parties.

Agentic AI systems' robustness and stability should be reliably assessed. Batch algorithmic tests might be not useful since a wide range of intended situations need assessing. Opponents would also need to be predicted. How would systems with continuous updates and developmental space be reliably assessed? Existing practices around AI might change, such as in-service red-teaming. A continuous reliable assessment approach would need developing, responding to outcomes from AI's inherent randomness and rapid adaptation.

8. Ethical Considerations

Recent technological strides have allowed for significant improvements in the risk management and fraud detection capabilities of financial institutions. These improved methods rely heavily on machine learning algorithms for their applications; however, they are by no means perfect. Towards the end of the previous decade, a new age of AI was ushered in with the advent of Transfer learning, Generative Models and Self-supervised learning techniques that greatly expanded the capabilities of machine learning. These techniques have been largely ignored in risk and fraud detection applications, and their implementation would provide great improvements over current systems. Furthermore, with these improvements in technology, new ethical concerns arise with which firms must engage. Agentic AI has great potential in changing the landscape of risk management and fraud detection. However, the inherent risks associated with the deployment of these systems warrants careful design and implementation.

Automation of processes has lengthened the decision chains involved in risk monitoring. This results in risk circumstances that may not have been previously foreseen and ones that may be deemed to be too complex for humans alone to respond to, even at the level of first response. Thus, the ultimate risk exposure is in the hands of machines. This exposes the firm to risks that are exacerbated by delays in the decision making time. These considerations must be taken into account when delegating parts of the risk control and detection processes to machines. Design decisions, as well as new practices in the financial firm, must be made so that true agentic functioning can occur. Moreover, AI systems are at risk of being captured. However, new frameworks for testing the robustness of AI systems against malintent are being developed. These are an example of the new novel solutions for newly arising risks that should be prioritized.

Firms should engage in interdisciplinary projects with academia to elaborate upon such solutions and to keep abreast of the state of the art in technical methods for risk systems. The risk estimation process in banks is imperfect and uncertainty neutral. AI systems provide a route towards newer methods of modelling risk contracts with higher weight on risky outcomes and further refinement of the models. As a side effect of the grander objectives of improving the modeling of risk and outcomes, a risk estimate could usually be bound, a practice not common in current systems. These estimates can help firm executives in parsing the operations of banks so as to reduce the model input space which can be notoriously difficult for humans to interpret. A multi-teamed approach should be taken towards implementation, where the risks associated with policy contracts are worked through on a systematic basis with all stakeholders of the contracts as a form of counterfactual analysis and stress-testing. Model uncertainty assessments should then be formulated based on this analysis. Having been a persistent sore-spot for the finance world for over a decade now since the inception of "black box" models, prospectively bound estimates ought to be available for all measurables.

8.1. Bias and Fairness in AI Models : As investment and changes in the regulatory landscape make machine learning (ML) methods exponentially more attractive, some risks arise. While there are many potential benefits, the view from above suggests the potential for misuse. The second half of this chapter will explore both sides of the same coin. Like any other widespread technology, greater control over the banking system and retail has come at both sides. If they bring uncharted capabilities to solve massive problems in the service of profitability, efficiency, and customer service, they also open doors for collusion, spying, disruptive growth of one-sided control of the financial system, suppression of internal competition, financial crises without precedent, and social neglect under exclusion from services.

Usage of biased patterns to learn to make predictions without accounting for possible underlying prejudices can lead to decisions that disproportionately harm certain social groups, such as the ill people left out of health services and credit. Financial services are no exception from this, with multiple works in the field warning against potential discrimination. Then, with the increasing prominence of ML in the aforementioned domains, its potential to exacerbate existing social inequities has been a reason of growing concern. Financial services are no exception from this, with multiple works in the field warning against potential discrimination. Systems can learn to make predictions exhibiting discrimination, but fixable social inequities in the ongoing world of covariates, such as health issues, language barriers, job supply/demand mismatch, and poverty are complicated to learn or account for in a more considerate way.

The goal of building systems that incorporate these concerns has given rise to the field of Fair ML. However, despite growing interest in the field, the state of the art relies on methods that predominantly focus on devising ways to measure unfairness and to mitigate it in algorithmic prediction tasks. Mitigation can be undertaken broadly by means of three approaches: pre-processing, in-processing, and post-processing. Whereas pre-processing assumes that the cause of

unfairness is bias in the data, in- and post-processing shift the onus to modeling choices and criteria. Research seems to be divided along the same lines in what concerns uncovering the source of bias in the ML pipeline.

8.2. Transparency and Accountability : Translucence addressing XAI's demands is often one of the fundamental questions relating to the public's perception of AI systems. Several players could consider how transparency could be achieved for a bank that plans to deploy agentic AI for risk management or fraud detection. On the one hand, the question of how decisions made by an agentic AI system should be communicated to its users and those affected is mainly relevant within the banking organization. On the other hand, the need for openness towards customers, the broader public, and governments is mainly a societal concern. These are typically questions for which the public has not provided clear answers yet. Which level of transparency should be sought, and which costs of achieving this kind of transparency does society consider acceptable? It seems that in the first group, banks should make provision for the eventuality that effective counter-explanations will be philosophically persuasive—a future as dark as any dystopian fable of anthropomorphic automata taking over control. Consequently, following the principle of least surprise, any information that an AI system could reasonably be expected to provide its users directly should therefore be made proactive. Banks should design agentic AI systems that are, at least for most actions, able to answer the very last question of Transparency for whom?, Who ought to see which level of AI information?. Beyond strictly rational requirements, banks should lay out paths of intervention for regulators, customers, and oversight processes. Central questions in this, apart from their actual societal setting, are: What skin do banks in the game have? What are the leadership roles? This is really pushing for a more operational level and more criteria such as: Specific timelines, probabilistic levels of compliance, incentive structures, acceptable penalties, and so on. Besides providing such detailed commitments about different parties such as users, corporate stakeholders, and regulators, banks should explore the window of opportunities. For instance, they could consider approaches that stimulate both the bottom-up processes that analyze information risk categories and the top-down processes for normative concepts of transparency regarding agentic AI systems. Banks must be extra vigilant about this, as societal trust could be more easily lost than built.

9. Implementation Challenges

While agentic AI has the potential to revolutionize risk management and fraud detection in banks, its implementation comes with a unique set of challenges. Banks must ensure model accuracy, security, and governance while also allowing agentic capabilities. This requires adaptive AI systems that can operate under human oversight and accept responsibility for their actions. However, it is difficult to maintain the agentic element of AI while also having adequate controls. Sustainability, data protection, and lending bias are all issues that a human monitor cannot be directly aware of. Similarly, fairness or accountability without autonomous agentic behaviour will only achieve narrow grounds. Such issues are still present in traditional AI systems, but they are

exacerbated through agentic capabilities. Individuals and organizations rely on complex models and algorithms for critical operations, but there are misunderstandings about how they function. This fear is magnified with agentic systems, which operate outside of a centralized control environment and could lead to significant economic losses.

Agentic AI and machine learning systems operate in a distributed fashion. There is a growing concern that the inner workings of these systems would be poorly understood, and thus that system behaviour could deviate significantly from intent. Ideally, an AI system would be simple and interpretable enough that any unintended behaviour could be foreseen. Presentation and assertiveness of behaviours could communicate these behaviours to users, such that any untoward consequences could be recognized early. AI agents that operate subconsciously and in opaque fashions could give rise to risks of behaviour that was unforeseen and untenable.

An engineering solution to these issues is a high level of adherence to Simplicity Understanding Transparency Explicability Communications (SUTEC) principles for the design of agentic systems. Exceptions to this principle based on aspects of deep learning performance could make it significantly less interpretable and comprehensible beyond sufficient accuracy or applicability. Moreover, representational hierarchies to support high level verification could drastically sacrifice performance gains. Like providing an external stimulus for understanding consciousness, too much augmenting of human intention in understanding AI could fail.

9.1. Technical Barriers : There are several technical barriers in the way of agentic AI. First of all, the issue of scaling up. ChatGPT is a very large model with hundreds of billions of parameters, trained on vast amounts of text, which together with the architecture decisions feeds into the emergent capabilities. It has been observed that, with larger models and more data, LLMs tend to exhibit a wider range of competencies. While the rise of the current models hints at scaling much wider levels of spark of intelligence, current understanding of these models renders them quite opaque and has not yet progressed to the theoretical ground where scaling laws could be properly exploited. For a bank to leverage agentic AI in its risk management work, an important open question is whether off-the-shelf (e.g. proprietary models) suffice, or if domain-specific effort hosted in-house would be a necessary price to pay for cross-application magical benefits.

The second important barrier is that of knowledge transfer. While there are several successful use cases with LLMs in bridging worlds of knowledge , there are strict limitations in most domains where “blind use” cannot be expected to work. Thus, inquiry into how to go about assembling and representing domain knowledge is of utmost importance. For agentic AI to work in governance of risk, the matter requires further exploration into how to design the interaction between the various AI agents to ensure proper refinement and checking of generated content. Currently there may be a rather naive division into a “dedicated risk sourcing agent” and a “check / advise / detect agent”. Both sides have malformations on their own from which comes the question to what extent it can be run in parallel and how assertions can be correctly matched to risks or questions.

Important questions of safety remain. There are a lot of accountability issues that follow from the complexity of the intelligence that is imprinted into LLMs as sovereign models. Despite ChatGPT and others being cute playhouses on the surface, there is a whole world of technical turf underneath which could hack into, gravely misuse, or bend the action of "computers in general". Complete predictive models could be used to detect any frailness in individual commercial banks, or countries, making for a "shadow bank" that is beyond control of all that make regulations or political (and non-political) invoices that are not AI-augmented.

9.2. Organizational Resistance: An organization's culture, which encompasses employee beliefs and behaviours both consciously and unconsciously, affects data management decisions and affects the risk management decisions banks make. Consequently, banks with a culture that prioritizes a conservative approach to decision-making in the face of risk (risk-averse) face greater risks. These banks would need to pool considerable data resources and wealth of knowledge that are abundant in intelligent transactions if they were to deploy Agentic AI effectively. Should they reveal sensitive information relating to these transactions, which would need to happen at least in part to achieve the required scale? Even if banks agree to new procedures and measures to keep data safe and in-house, education and training will need to get new staff on board. Both to develop the existing systems to integrate with Agentic AI and to extract assets and outputs from these external systems, this would be extremely resource- and time-expensive. There may be members of staff with the relevant experience, but banks may need to hire in an entirely new team of experts (which would involve searching for prospective hires then training them in what the existing systems are). Organizations are often resistant to change due to individuals' concern about their livelihoods. Abolishing roles entirely may not even be considered here; exploiting Agentic AI would likely require staff to operate the systems and oversight roles to vet the recommendations and decision-making. However, despite Agentic AI's promise, these roles and responsibilities would need considerable reevaluation after it was integrated. Agentic AI would undoubtedly transform these operations and potentially reduce staffing requirements over time. Staff might find it in their interests to delay deployment whilst they continue to exercise their experience and domain knowledge. The suggestion of Agentic AI disrupting risk management and fraud detection practices would challenge defenders of the status quo and attract concern over a significant intellectual shortfall and, by extension, an existential threat in the face of highly advanced AI systems. Even if those individuals affecting change could be identified and won over, perhaps with the assistance of human AI, concerns over job security might transfer to executive think tanks as they tried to circumnavigate these systems. Staff would need to be assured that Agentic AI would not undermine them or ultimately replace them. It is conceivable that bespoke systems would emerge to similarly disrupt the agents, accrue data faster than they could be parsed, or produce bad-faith outputs that obfuscate attempts to peer deep into the thought process.

Equ 3: Agentic AI Utility Optimization Model

$$\max_{\pi} \mathbb{E}_{s_t, a_t \sim \pi} \left[\sum_{t=0}^{T} \gamma^t \left(R(s_t, a_t) - \lambda \cdot R_f(s_t, a_t) \right) \right]$$

- π: Agentic AI policy (decision-making strategy)
- s_t, a_t: State and action at time t
- γ: Discount factor (future reward weighting)
- $R(s_t, a_t)$: Reward from successful risk mitigation
- $R_f(s_t, a_t)$: Penalty/cost from fraud events
- λ: Fraud risk weighting parameter

10. Future Trends in Agentic AI

Intelligent financial agents have potential in the area of risk management and fraud detection in banks. In many systems today, processes related to the business of risk management and fraud detection in banks are performed either by software alone or by humans and software. Risk management, in conjunction with supervisory regulators, monitors the level of risk exposure of credit and market risks, monitors the adequacy of liquidity and capital, determines and implements policies to limit risk exposures, and implements stress testing and scenario modelling. Fraud detection continuously monitors for fraud using knowledge of historical and statistical behaviours and patterns of accounts, transactions, and their variables, continually assesses their likely goods and services usage, and raises alerts for potentially fraudulent behaviours and patterns who in turn implements investigations.

Many of the decisions made in risk management rely on financial stakeholders' and business regulators' opinions, and also rely on explanations from data scientists, traders, and analysts, while fraud detection is often performed collaboratively between software agents and human financial forensic investigators. In such systems, human reasoning about the business and regulatory policies to execute, human explanations of a prevention or alert in the light of knowledge of how the businesses, regulators, and data scientists intended, and human forensic investigations of a fraud alert could use a more systematic, performance methodical, holistic, reliable, and reliable approach. Agent-based, knowledge-enabled systems could model how such systems might function, and the theory might explicitly explain the general reasons for agentic AI systems and explain the credit derivatives risk management process in a wider range of business and regulatory settings, encompassing business purpose, analysis of exposure, bid and ask scenarios and prices, trader price changes, portfolio valuation, capital charge, stress testing, and scenario generation of extreme rated events.Future Trends in Agentic AI focus on research questions that further explore the potential of agentic AI in decision systems for financial services, and which elaborate in more detail the many challenges and opportunities in those research questions.

Fig 5: Future Trends in Agentic AI.

10.1. Predicted Developments in AI Technology: As discussed above, agentic AI will certainly revolutionise the entire financial services industry. This includes both the future developments of AI technology as well as the expected challenges banks will likely face when deploying agentic AI. After explaining the important and pervasive developments in AI introduced in the previous sections, the focus will shift to how these developments are very likely to impact banking, compliance and regulatory efforts, particularly involving market and credit risk. Furthermore, the limitations of auto-documentation, and the likely legal and societal challenges regarding risk generation, captioning, litigious process management, "inner loop bias," and agent accountability will be discussed. This section will also draw empirical evidence from the efficiency gain from automating model development and other critical aspects of implementation, maintaining, monitoring and updating models and their economic impacts, particularly with respect to regulatory expectations and challenges for banks and opportunities for regulators.

AI would no doubt transform the financial services landscape. As such, and particularly because of banks' systemic importance, investor protection, volatility and contagion implications, it would be useful to scrutinise plausible benefits and risks as well as approaches to harnessing this new technological frontier, with the focus being the area of compliance and risk management. AI would take care of regulatory compliance and monitoring, as well as risk measurement, market risk, portfolio and counterparty exposure management, liquidity and profitability stress testing, model development, and model governance. This initial survey covers a flavour of the possibilities arising from non-human agents in tandem with advanced AI, highlighting important challenges and concerns, legal and ethical as well as machine governance issues, agent insight and understanding issues, as well as societal and regulatory risks stemming from agent use. Domain-specific and hierarchical structures were proposed to partition contextual boundaries. Future work would explore banks' agents and flesh out the technical aspects of this immense opportunity.

10.2. Shifts in Banking Practices: This section reflects on changes to established banking practices due to Contextual Agentic AI (CAAI). As CAAI advances, banks will have to look at developing terms of service and proof-of-concept testing to comply and adapt, particularly around custodied AI. In developing proof-of-concept CAAI, banks should consider how to restrict agentic evolution or behaviour. This begs the question, what recording of agentic endeavours should banks consider disclosing beyond reasoning logs? CAAI will likely co-evolve and adapt to the emergence

of a custodied agentic AI market, leading to an arms race among fin-tech directors to acquire and deploy the best agents. To remain competitive, banks must evaluate the degree of monitoring and custodial investment in pre-deployed agents to set up a pool of agents of higher value than their parent-developing institution. Banks need robust tests to prove that their expensive CAAIs can withstand high-stakes reasoning tests before deployment.

11. Conclusion

As financial markets become increasingly connected, the potential for systemic risk to increase and amplify is becoming more pronounced. Emerging financial products and business processes have hitherto unknown risks that may be beyond the understanding of any one regulator, and rapid response becomes paramount in the event of emerging distress or disaster. Machine learning models functioning as agentic AI can provide regulators with great aid in monitoring and measuring such risks, and offer actionable suggestions for mitigating the consequences but may also create new systemic risk via unexpected consequences resulting from powerful capabilities. More broadly, as advancements in AI technology lead to the potential for agents with human-like capabilities, the risks associated with their misalignment may become pronounced. In addition to high-stake regulation of agentic AI, more mundane regulatory efforts such as traditional risk management or compliance systems may need to be fundamentally rethought. Systems can be built to monitor and drive into compliance agentic AI functioning in the more mundane cases of corporate automation. By relaxing the requirements for performance or frugally narrowing the space of possible actions, regulators can control agentic AI while still allowing it to function effectively and reinforce familiar hierarchically-monitored systems.

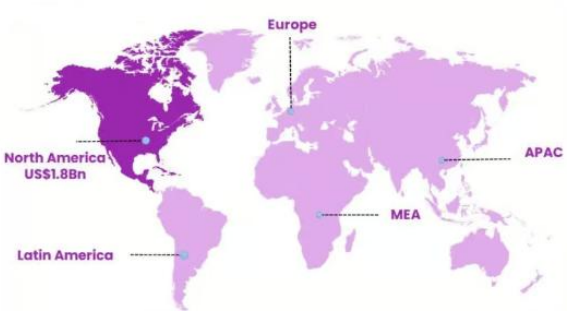

Fig 6: Agentic AI in Risk Management and Fraud Detection in Banks.

However, even mundane agentic AI poses new challenges for regulation. Critical performance indicators of agentic AI may inherently lend themselves to manipulation, likely creating a regulatory arms race between financial companies and regulators. Safeguards implemented to reduce the risk of misalignment can themselves be manipulated, or bypassed altogether. The scale and speed of modern finance, and comparable projects in other domains, is also expected to

continuously increase. Such a pace of change combining previously disparate systems and settings, as it has in academia, may create social and economic blind spots that regulators cannot simply react to. Even more broadly, regulation of safety in deployment or behavior cannot meaningfully occur across the many separate domains that agentic AI may control over time or be deployed across. Decision making outside the training distribution of an agentic AI is notoriously difficult to regulate, making it infeasible to ensure that the safety mechanisms and internal reins would apply as intended across domain shifts. Reinforcement learning or deep learning applied in any of the many fields in which it has been adopted create new misalignment risks that are both novel and difficult to reason about. To face these new challenges, academia as a community is called upon to better understand the classes and potential of risk domains and create safe and accessible by design systems that are reliable in novel situations.

12. References

[1] Nuka, S. T., Chakilam, C., Chava, K., Suura, S. R., & Recharla, M. (2025). AI-Driven Drug Discovery: Transforming Neurological and Neurodegenerative Disease Treatment Through Bioinformatics and Genomic Research. American Journal of Psychiatric Rehabilitation, 28(1), 124-135.

[2] Annapareddy, V. N. (2025). The Intersection of Big Data, Cybersecurity, and ERP Systems: A Deep Learning Perspective. Journal of Artificial Intelligence and Big Data Disciplines, 2(1), 45-53.

[3] Recharla, M., Chakilam, C., Kannan, S., Nuka, S. T., & Suura, S. R. (2025). Revolutionizing Healthcare with Generative AI: Enhancing Patient Care, Disease Research, and Early Intervention Strategies. American Journal of Psychiatric Rehabilitation, 28(1), 98-111

[4] Kumar, B. H., Nuka, S. T., Malempati, M., Sriram, H. K., Mashetty, S., & Kannan, S. (2025). Big Data in Cybersecurity: Enhancing Threat Detection with AI and ML. Metallurgical and Materials Engineering, 31(3), 12-20.

[5] Chava, K. . (2025). Dynamic Neural Architectures and AI-Augmented Platforms for Personalized Direct-to-Practitioner Healthcare Engagements. Journal of Neonatal Surgery, 14(4S), 501–510. https://doi.org/10.52783/jns.v14.1824.

[6] Manikandan, K., Pamisetty, V., Challa, S. R., Komaragiri, V. B., Challa, K., & Chava, K. (2025). Scalability and Efficiency in Distributed Big Data Architectures: A Comparative Study. Metallurgical and Materials Engineering, 31(3), 40-49.

[7] Suura, S. R. (2025). Integrating genomic medicine and artificial intelligence for early and targeted health interventions. European Advanced Journal for Emerging Technologies (EAJET)-p-ISSN 3050-9734 en e-ISSN 3050-9742, 2(1).

[8] Chabok Pour, J., Kalisetty, S., Malempati, M., Challa, K., Mandala, V., Kumar, B., & Azamathulla, H. M. (2025). Integrating Hydrological and Hydraulic Approaches for Adaptive Environmental Flow Management: A Multi-Method Approach for Adaptive River Management in Semi-Arid Regions. Water, 17(7), 926.

[9] Burugulla, J. K. R. (2025). Enhancing Credit and Charge Card Risk Assessment Through Generative AI and Big Data Analytics: A Novel Approach to Fraud Detection and Consumer Spending Patterns. Cuestiones de Fisioterapia, 54(4), 964-972.

[10] Peruthambi, V., Pandiri, L., Kaulwar, P. K., Koppolu, H. K. R., Adusupalli, B., & Pamisetty, A. (2025). Big Data-Driven Predictive Maintenance for Industrial IoT (IIoT) Systems. Metallurgical and Materials Engineering, 31(3), 21-30.

[11] Recharla, M., Chakilam, C., Kannan, S., Nuka, S. T., & Suura, S. R. (2025). Harnessing AI and Machine Learning for Precision Medicine: Advancements in Genomic Research, Disease Detection, and Personalized Healthcare. American Journal of Psychiatric Rehabilitation, 28(1), 112-123.

[12] Kumar, S. S., Singireddy, S., Nanan, B. P., Recharla, M., Gadi, A. L., & Paleti, S. (2025). Optimizing Edge Computing for Big Data Processing in Smart Cities. Metallurgical and Materials Engineering, 31(3), 31-39.

[13] Kannan, S. (2025). Transforming Community Engagement with Generative AI: Harnessing Machine Learning and Neural Networks for Hunger Alleviation and Global Food Security. Cuestiones de Fisioterapia, 54(4), 953-963.

[14] Sriram, H. K. (2025). Leveraging artificial intelligence and machine learning for next-generation credit risk assessment models. European Advanced Journal for Science & Engineering (EAJSE)-p-ISSN 3050-9696 en e-ISSN 3050-970X, 2(1).

[15] Chakilam, C., & Rani, P. S. Designing AI-Powered Neural Networks for Real-Time Insurance Benefit Analysis and Financial Assistance Optimization in Healthcare Services.

[16] Chakilam, C., Kannan, S., Recharla, M., Suura, S. R., & Nuka, S. T. (2025). The Impact of Big Data and Cloud Computing on Genetic Testing and Reproductive Health Management. American Journal of Psychiatric Rehabilitation, 28(1), 62-72.

[17] Suura, S. R. (2025). Integrating Artificial Intelligence, Machine Learning, and Big Data with Genetic Testing and Genomic Medicine to Enable Earlier, Personalized Health Interventions. Deep Science Publishing

[18] Kumar Kaulwar, P. (2025). Enhancing ERP Systems with Big Data Analytics and AI-Driven Cybersecurity Mechanisms. Journal of Artificial Intelligence and Big Data Disciplines, 2(1), 27-35.

[19] Suura, S. R. (2025). Agentic AI Systems in Organ Health Management: Early Detection of Rejection in Transplant Patients. Journal of Neonatal Surgery, 14(4s).

[20] Dodda, A., Polineni, T. N. S., Yasmeen, Z., Vankayalapati, R. K., & Ganti, V. K. A. T. (2025, January). Inclusive and Transparent Loan Prediction: A Cost-Sensitive Stacking Model for Financial Analytics. In 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI) (pp. 749-754)..

[22] Challa, S. R. The Intersection of Estate Planning and Financial Technology: Innovations in Trust Administration and Wealth Transfer Strategies. GLOBAL PEN PRESS UK.

[23] Nuka, S. T. (2025). Leveraging AI and Generative AI for Medical Device Innovation: Enhancing Custom Product Development and Patient Specific Solutions. Journal of Neonatal Surgery, 14(4s).

[24] Annapareddy, V. N. (2025). Connected Intelligence: Transforming Education and Energy with Big Data, Cloud Connectors, and Artificial Intelligence. Deep Science Publishing.

[25] Mashetty, S. (2025). Securitizing Shelter: Technology-Driven Insights into Single-Family Mortgage Financing and Affordable Housing Initiatives. Deep Science Publishing.

[26] Sriram, H. K. (2025). Generative AI and Neural Networks in Human Resource Management: Transforming Payroll, Workforce Insights, and Digital Employee Payments through AI Innovations. Advances in Consumer Research, 2(1).

[27] Challa, K., Chava, K., Danda, R. R., & Kannan, S. EXPLORING AGENTIC AI Pioneering the Next Frontier in Autonomous DecisionMaking and Machine Learning Applications. SADGURU PUBLICATIONS.

[28] Challa, S. R. (2025). Advancements in Digital Brokerage and Algorithmic Trading: The Evolution of Investment Platforms in a Data Driven Financial Ecosystem. Advances in Consumer Research, 2(1).

[29] Ganti, S., Vankayalapati, R. K., Krishnamoorthy, P., Thakare, P. S., Nayak, U. A., & Vignesh, P. (2025, February). Enhancing IoT-Driven Smart Home Security and Automation with a GCN Model. In 2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS) (pp. 1-6). IEEE.

[30] Syed, S., Nampalli, R. C. R., Nikam, M., Krishnan, T., & Perada, A. (2025, February). IoT-Driven Environmental Pollution Monitoring with a Deep Attentional Hybrid Transformer Model. In 2025 International Conference on Emerging Systems and Intelligent Computing (ESIC) (pp. 356-361). IEEE.

[31] Nampalli, R. C. R., Syed, S., Bansal, A., Vankayalapati, R. K., & Danda, R. R. (2024, December). Optimizing Automotive Manufacturing Supply Chains with Linear Support Vector

Machines. In 2024 9th International Conference on Communication and Electronics Systems (ICCES) (pp. 574-579). IEEE.