Intelligent & Human Futures
WISDOM ACADEMIC https://doi.org/10.63808/ihf.v2i4.477 Volume 2 , Issue 4

Literature Review

Artificial Intelligence in Investment Banking Risk Management: Applications, Benefits, and Governance Challenges

Ratul Hasan[1,*]
[1]Department of Management Information Systems, Begum Rokeya University, Rangpur, Bangladesh Rangpur-5404, Bangladesh
[*]**Correspondence:** Ratul Hasan, ratul.12222042@student.brur.ac.bd

© The Author(s) 2026.
This article is published by Wisdom Academic Press Ltd. under a Creative Commons Attribution 4.0 International License. The license permits use, sharing, adaptation, distribution, and reproduction in any medium or format, provided that appropriate credit is given to the original author(s) and the source, a link to the license is supplied, and any modifications are indicated. Unless otherwise stated, all images and third-party materials included in this article are also covered by the same license. For any material not covered by this license, if the intended use is neither permitted by applicable law nor allowed under the license terms, direct permission must be obtained from the copyright holder. To view a copy of this license, please visit http://creativecommons.org/licenses/by/4.0/.

Abstract

Artificial Intelligence (AI) is revolutionizing the investment banking landscape and its impact on risk management is undeniable, with the ability to analyze vast volumes of complex and heterogeneous data to make more informed and timely decisions regarding market risk, credit risk, operational risk, and compliance risk. The design of this study is qualitative, descriptive, and secondary data analysis, based solely on academic literature that discusses AI applications, benefits, and challenges in investment-banking risk management, from which the synthesis of how AI tools are used in the different stages of the risk management process (identification, measurement, monitoring, and mitigation) is taken. Thematic content analysis is applied to identify the key areas of the common applications of AI: short-term market forecasts, AI-enhanced credit scoring, anomaly-based fraud detection, and natural language processing (NLP) for surveillance, and the perceived benefits and implementation challenges. The review indicates that AI can have a substantial impact on predictive accuracy, early identification of anomalies and unethical activity, operational efficiency, and integration of multiple data sources, which can add to the resilience of banks when implemented in a strong governance structure. Meanwhile, the growing complexity of models, data quality challenges, lack of model explainability, potential model bias, and shifting regulatory expectations add a new layer of model and governance risk that needs to be managed. The study concludes that AI should be considered a useful complementary tool, not a replacement, for the traditional quantitative models and investment bankers' experience, and recommends that investment banks have clear AI strategies, improve model risk management, invest in cross-functional capabilities, and integrate ethical and regulatory aspects throughout the AI lifecycle.

Keywords

Artificial intelligence, Investment banking, Risk management, Market risk, Credit risk, Operational risk, Compliance risk, Machine learning, Deep learning, Model risk governance, Explainable AI

1. Introduction

1.1. Background of the Study

One of the most rapidly changing trends in the international financial services industry has been the rapid evolution of artificial intelligence (AI). While the requirement for the use of very advanced technologies has been dominant in investment banking where the products are complex, transactions are high, and stakes are high, to stay competitive and abide by increasingly stringent regulatory standards. The industries traditional methods of measuring and controlling investment risk exposures have been statistical models, analysis of historical data and the exercise of judgment. These methods work well in certain situations, but make assumptions about consistent relationships, commonly distributed returns and slowly changing environments that are generally incorrect when environments are stressed or undergoing structural change.

However, some of these limitations can be overcome with the help of AI technologies, like machine learning, deep learning, and natural language processing, which can learn more complex patterns directly from the data, adapt to new information, and process large amounts of structured and unstructured data. For example, AI models can uncover nonlinear relationships between risk factors that are difficult to identify, indicate obscure communities of high-risk counterparties, or identify anomalous trading or communication patterns that may indicate fraud or market abuse. At the same time, the implementation of AI in risk management creates new challenges associated with model transparency, data governance, and ethical use (Habbal et al., 2024). From these changes, the role of AI in the innovation of risk management in investment banking, and its implications for risk management in general, are relevant subjects of discussion between scholars, practitioners and regulators.

1.2. Problem Statement

While the future of AI is bright, the majority of investment banks are still using legacy IT systems, siloed data sources and risk models that were designed for previous market cycles. These models might lack the flexibility to manage rapidly evolving market dynamics, interdependent risk exposures, and novel risk profiles associated with new business models and compliance requirements that can arise in a digital, data-intensive business landscape. Other institutions keep risk data in silos, risk models are only calibrated on a periodic basis and key monitoring and reporting processes are still manual. That could leave banks vulnerable to new risks when it is too late to discover them, or to

recognize the loss they may sustain, because of the risks identified.

Implementing AI in such environments is not as simple as it sounds. There are many different types of AI models, and they may be difficult to interpret, be highly dependent on the quality of the data they were trained on, have implicit biases, and behave unpredictably in new or difficult situations. So, it's up to the regulators and senior management to do two things – enable innovation and maintain transparency, strength, and alignment with regulatory and ethical criteria for their tools. The need for a systematic consideration of the use of AI in the existing investment banking environment for the management of risk, and the benefit of AI, is very strong, alongside creating new risks and governance issues. In response to these needs, the current report is a synthesis of literature and implications for practice are drawn from the literature.

1.3. Objectives of the Study

RO1: To analyze the existing and emerging uses of AI in the identification, measurement, and monitoring of major risks market, credit, operational, and compliance in investment banks.

RO2: To investigate the possible advantages of AI-based tools in improving the accuracy, timeliness, and efficiency of risk management processes.

RO3: To determine the key challenges and risks related to the implementation of AI, such as model risk, data quality problems, ethical issues, and regulatory limitations.

1.4. Scope of the Study

The scope of this report is the use of AI in risk management activities by investment banks, as opposed to the wider sphere of retail banking or insurance (Ashta & Herrmann, 2021). Investment banking activities that involve high levels of market and credit exposures as well as operational and compliance risk include trading, structuring of derivatives, security underwriting and advisory. The paper will touch on applications of AI to these risk functions, such as AI-based market-risk models, AI-assisted credit scoring and counterparty risk evaluation, operational-loss anomaly detection, and AI-based compliance and conduct-risk surveillance, to mention a few.

No survey, interview or direct access to the bank records was done to collect primary data, the analysis is based purely on secondary data with no primary data reported throughout this study. The study is conceptual and descriptive with emphasis on synthesizing existing academic literature to the best of the researcher's ability. Geographically, the report has a global focus, with examples of key financial hubs and recognizing that there may be variations in the practice and application of regulation and technology in different jurisdictions. It focuses on advances within the past decade, during which AI uses in finance started to gain vital traction (Lu, 2019).

1.5. Organization of the Report

The report is broken down into major sections for readability. Following this introduction, Chapter 2 provides a literature review and background, covering the theories relevant to this topic, the definition of key concepts, discussions of empirical studies about the application of AI in the field of risk management, and a conceptual framework that connects the AI tools with the various stages of the risk management process. The methodology of the study is explained in chapter 3, research design, research sources, sampling, data collection and analysis and limitations of the chosen methodology. The discussion and analysis of the above is provided in Chapter 4, which outlines the market, credit, operational and compliance risks applications of AI and provides a summarized overview of common AI applications, benefits and challenges. Chapter 5 brings the report to an end by summarizing important themes from the review, overall conclusions, practitioner recommendations and suggestions for future research.

2. Literature Review / Background

2.1. Theoretical Background

The management of risk in investment banking is grounded in several underlying theories and models in finance (Ho & Lee, 2004). Modern portfolio theory focuses on diversification and the risk-return trade-off, indicating that an investor can optimize portfolios by weighing expected returns against the variance of returns. Building on this, the sensitivity of portfolios to market movements has traditionally been measured using risk measures such as variance, covariance, and beta. Value-at-Risk (VaR) models, which approximate the potential loss that might occur over a specified time frame at a specified confidence level, have become common instruments for measuring market risk and are embedded in regulatory capital requirements.

Such conventional models, however, typically assume normally distributed returns, linear relationships among variables, and stable historical patterns, assumptions that are less likely to hold during periods of stress or when dealing with complex financial instruments. These frameworks are challenged and extended by AI-based models, especially machine learning and deep learning, which allow non-parametric, data-driven modeling capable of capturing nonlinear interactions, structural breaks, and multidimensional feature spaces. Theoretically, AI does not replace existing risk theories but introduces new instruments for estimating and managing risk in these broader contexts (Giudici et al., 2024). At the same time, the opaque nature of many AI models can undermine interpretability, introduce model risk, and limit the capacity of human decision makers to fully understand and manage AI-driven risk processes.

2.2. Key Concepts and Definitions

In the context of financial services, AI can be described as any computational system which

exhibits the same or greater capability as a human intelligence system in one or more of the following aspects: Ability to learn from data, ability to recognize patterns, ability to predict new patterns, ability to recommend new patterns. In AI, machine learning is a branch of algorithms that become more proficient with more data, without the need for them to be specifically programmed for every possible situation (Sarker, 2021). Supervised learning techniques (linear and logistic regression, decision trees, random forest, gradient boosting, neural networks) learn from labeled data while unsupervised techniques (clustering, anomaly detection) learn from unlabeled data. Deep learning (also known as multi-layered learning) is a type of machine learning that employs a multi-layered neural network to detect patterns or relationships that are more complex, particularly in high-dimensional data such as images or text.

There are multiple types of risk that are important for investment banking risk management. Market risk is the risk of losses from changes in prices in the market, such as equity prices, interest rates, foreign exchange rates and commodity prices. Credit risk is the risk of losing an amount because the counterparty fails to fulfil its financial obligations (Gregory, 2012). Operational risk refers to losses resulting from inadequate or failed internal processes, systems and people, or from external events like fraud and cyber incidents. Regulatory/compliance risk (may be combined with conduct and legal risk) is the risk of sanctions, monetary loss or damage to reputation due to failure to adhere to the law or company policy. Model risk is a cross-cutting concept that refers to any risk that is caused by an incorrect or misused model, such as one using artificial intelligence.

2.3. Empirical Studies on AI in Risk Management

Several academic and industry studies have cited the increasing adoption of AI tools in risk management roles within large financial institutions. Global investment banks are turning to machine learning models for short term prediction, intraday VaR, scenario generation and more, and some of these banks are finding that the machine learning models are more effective than linear models, according to consulting firms and industry associations. A similar line of academic research has demonstrated that machine learning systems like random forests, gradient boosting machines, and deep neural networks can enhance predictions of extreme market returns and losses, especially when fed with dense data on markets and macroeconomic data.

For credit risk, research shows that AI-driven credit scoring models can use more data to better predict defaults, such as financial ratios, market indicators and alternative data like news sentiment. AI's use in operational and compliance risk research has proven to be effective in detecting anomalies in transaction data, identifying suspicious patterns that could signal fraud or money laundering, and spotting signs of potential misconduct in unstructured data like emails, chat logs, and recorded conversations (Attipalli et al., 2025). However, it also warns of the limitations of AI models, including

reliance on data quality, overfitting, and the lack of validation and interpretability, all of which are of concern to regulators and internal risk committees. These findings from the reviewed studies underscore the importance of strong governance structures when adopting AI.

This picture is expanded and deepened by the more recent and specialist literature, and by adding to the evidence base considered in this report. As for explainability and fairness, this systematic literature review of 43 peer-reviewed papers on explainability and fairness in AI-based credit scoring reveals that the metrics used to measure explainability and fairness are mainly developed separately from one another, there is a minimal and unquantified trade-off between explainability and performance, and fairness constraints exhibit a more consistent and tunable trade-off with accuracy, where applicable (Bahlool et al., 2026). The governance themes highlighted throughout this report are further highlighted by regulatory instruments that are placing new demands on the human-in-the-loop oversight and adverse-action justification of algorithmic credit decisions, as outlined in the U.S. Equal Credit Opportunity Act, Fair Credit Reporting Act, the EU AI Act, and similar guidance from the ASEAN area. In operational and compliance risk, recent research on the use of explainable-AI modules for high-risk client flagging in transactional and know-your-customer data shows how supervised machine learning models can be applied to detect high-risk clients and cautions against the ongoing reliance on data quality and labeled instances of proven misconduct (Namdar et al., 2025). Financial-sector authorities, at the Financial Stability Board level, identified model risk, data quality issues, dependency on third parties and vendors, and governance gaps as key vulnerabilities stemming from the increasing use of AI in the financial sector, in addition to operational efficiency and analytics benefits (FSB, 2024). They are used directly in Sections 4.6, 4.8 and 4.9 below, to provide information to the governance, critical analysis and regulatory comparison discussion.

2.4. Conceptual Framework

This report builds on the theories and empirical findings presented above and employs a conceptual framework that connects the capabilities of AI to each of the stages in the investment banking risk management process. The framework outlines four main stages: risk identification, risk measurement, risk monitoring and risk mitigation. During the risk identification stage, AI systems can identify new risks using anomaly detection algorithms and NLP tools, which can detect abnormal trading activities, counterparty behavior, or external information from news and social media. During the risk measurement stage, ML and deep learning algorithms can enhance the estimation of potential losses, default probabilities, and exposure distributions by uncovering intricate relationships from historical and real-time data.

During the risk monitoring process, AI can help banks monitor shifts in risk profiles and identify red flags that occur across markets, portfolios, and internal communications, faster than it can be done

manually or through rules. Lastly, during the risk mitigation stage, AI can assist with decision making by simulating scenarios of stress, assessing the possibility of hedging, and suggesting measures to take across different market conditions. The framework highlights the need for model governance, validation, and ethical supervision throughout, as they are all key to the management of model risk, data risk, and conduct risk in relation to AI. The structure of this framework will be followed in the discussion chapter, where specific types of AI applications are associated with risk types and risk management stages.

2.5. Positioning and Contribution Relative to Existing Reviews

Several existing reviews have already established that AI can improve pattern recognition and predictive capability in finance while introducing explainability, bias, data-quality, and governance concerns (e.g., Habbal et al., 2024; Sarker, 2021); this report does not treat that broad observation as its contribution. Instead, this report offers three narrower contributions. First, rather than surveying AI in banking or finance broadly, it synthesizes the literature specifically through the four-stage risk management process common to investment banking practice and regulatory expectation (identification, measurement, monitoring, and mitigation, as set out in the conceptual framework above), and it maps AI applications, benefits, and challenges onto each stage separately for market, credit, operational, and compliance risk rather than presenting them as an undifferentiated set of use cases. Second, it explicitly integrates the governance and regulatory literature (Section 4.6 and Section 4.9) with the risk-management literature, rather than discussing AI governance as a separate, generic topic; the report distinguishes conventional financial model risk from the additional risk categories that become material once machine-learning and deep-learning systems are deployed (e.g., data drift, adversarial vulnerability, and the performance-explainability trade-offs documented in recent credit-scoring reviews). Third, the report's critical-analysis section (Section 4.8) explicitly separates benefits that are empirically demonstrated in the reviewed studies from those that are reported by industry sources and those that remain theoretically anticipated, a distinction that is often blurred in narrower, single-risk-type reviews. The intended audience for these contributions is risk managers, model-risk and governance functions, and academic readers seeking a risk-process-oriented, governance-integrated synthesis rather than a purely technical survey of machine-learning methods.

3. Methodology

3.1. Research Design

This study adopts a qualitative, descriptive, narrative (conceptual) review design based on the thematic synthesis of secondary literature (Wacker, 1998). The narrative-review design is appropriate because the objective is to understand and explain how AI is being discussed and adopted in the risk

management literature relevant to investment banking, rather than to test a specific hypothesis against primary quantitative data. As a conceptual and narrative review, this study does not follow a systematic-review or scoping-review protocol (e.g., PRISMA); it does not report a formal search-and-screening flow diagram, and it does not involve statistical meta-analysis. This distinction is made explicit throughout the chapter, and the reporting of "findings" in this study refers to themes identified through the review of the literature rather than to results generated from original data collection or testing.

3.2. Data Sources

The study is based exclusively on secondary sources, specifically peer-reviewed academic journal articles, scholarly books, and academic conference or working papers addressing AI, quantitative finance, and risk management. No interviews, surveys, proprietary bank data, or other forms of primary data collection were undertaken, and no such primary data are used anywhere in this study, including in the discussion chapter. The complete list of sources drawn upon is provided in the reference list; all citations in this report are to items appearing in that list.

3.3. Population, Sample, and Search Process

The conceptual population of interest is the entire universe of investment banks (both global and regional) that are involved in capital markets and face meaningful market, credit, operational and compliance risk. Of the literature that exists on this population, a purposive sample of 15 sources was identified and cited throughout this report, with most of the literature published in the past five years (2019-2025) to reflect recent practice. Sources were identified by performing keyword searches for the term "artificial intelligence in investment banking" and "machine learning in market risk" (and the corresponding backward and forward citations) in the databases Google Scholar and SSRN (Martín-Martín et al., 2018) using the terms mentioned above.

Because this is a narrative rather than a systematic review, the number of records initially retrieved and subsequently excluded at each stage of the search was not formally logged, and no PRISMA-type flow count is reported; this is acknowledged as a limitation in Section 3.5. Sources were included where they (a) discussed an AI or machine learning technique or system, (b) related this discussion to investment banking, capital markets, or a directly comparable financial-risk context, and (c) were peer-reviewed, from a recognized academic publisher, or a reputable indexed working paper. Sources were excluded where their focus was unrelated to financial risk management (for example, purely technical AI research with no risk-management application) or where the connection to investment-banking risk management was only tangential. Relevance was judged by the author through review of titles, abstracts, and, where necessary, full texts, against these criteria.

3.4. Data Analysis Techniques

Qualitative thematic content analysis was used to analyze the selected sources, consistent with the overall research design (Vaismoradi et al., 2013). First, each source was read in the aim of identifying recurring themes relating to AI applications in different risk categories, perceived benefits (e.g., improved predictive accuracy or efficiency), and challenges (e.g., lack of explainability or data quality problems). Second, these initial themes were grouped into higher-level categories corresponding to the main risk types addressed in this report market, credit, operational, and compliance risk together with cross-cutting themes on model risk and governance. Third, the resulting themes were mapped onto the conceptual framework described in Section 2.4, distinguishing risk identification, measurement, monitoring, and mitigation, to analyze how AI is discussed in relation to each stage of the risk management process.

Coding was performed manually by the sole author of this study, without the use of qualitative-analysis software and without a second coder; consequently, no inter-coder agreement statistics are reported, and this is acknowledged as a limitation in Section 3.5. No formal quality-appraisal instrument (such as a critical-appraisal checklist) was applied to individual sources; instead, source credibility was assessed qualitatively based on peer-review status, publisher reputation, and indexing in recognized academic databases. To present the themes in an accessible way, Chapter 4 includes a summary table (Table 1) that synthesizes common AI applications, tools, benefits, and challenges by risk category, drawn from the reviewed literature.

3.5. Limitations of the Methodology

There are some limitations that need to be pointed out regarding the methodology chosen. First, the study is based on publicly available secondary sources only, so the discussion will not be as granular and specific as it could be because it does not incorporate any primary data collection and won't examine details of the confidential internal practices or proprietary AI models of individual investment banks. Second, this is a narrative review and the search and screening process was not documented in a reproducible flow diagram, and if the same review had been done by a different reviewer with the same keywords, slightly different sources may have been found; the review should therefore be treated as illustrative and a synthesis of review rather than a comprehensive review of the literature. Third, coding was done by only one person and there was no second coder or formal quality-appraisal instrument, thus, coding cannot be independently verified for inter-coder reliability. Fourth, AI and finance are fields with quick updates in technology and regulations, so some information may need updating frequently even with the most recent sources (Lee, 2020). Fifth, the qualitative and descriptive approach does not allow for quantitative estimates of the performance improvements or risk reductions that AI can deliver, nor does it enable to conduct statistical hypothesis

testing – any numbers used in Chapter 4 are presented as illustrative only.

4. Discussion and Analysis

4.1. Overview of Findings

Based on the reviewed literature, AI is discussed as being implemented across several risk management functions in investment banking, particularly market, credit, operational, and compliance risk. Literature generally presents AI tools as capable of processing large, complex datasets, identifying subtle patterns, and supporting faster, more informed risk decisions, while also introducing new categories of model, data, and governance risk that require careful management.

4.2. Synthesis of AI Applications by Risk Category

Table 1

AI applications in investment banking risk management (synthesized from the reviewed literature)

Risk type	Typical AI applications	Main AI tools/techniques	Key benefits	Main challenges
Market risk	Short-term price and volatility forecasting; intraday VaR; stress scenario generation	ML regression, deep learning (e.g., LSTM), ensembles	Better capture of nonlinear relations and tail risk; more timely risk measures	Model complexity; lack of transparency; overfitting; regulatory acceptance
Credit risk	Counterparty credit scoring; PD estimation; early-warning systems	Gradient boosting, random forests, neural networks, alternative-data analytics	More granular and dynamic credit assessment; earlier detection of deterioration	Data quality and availability; potential bias; explainability to regulators and clients
Operational risk	Fraud detection; incident pattern recognition; process monitoring	Anomaly detection, clustering, supervised ML on loss data	Earlier detection of unusual events; improved loss prevention	Incomplete incident data; evolving attack patterns; risk of missing novel failures

Compliance risk	Trade and communication surveillance; AML/CTF monitoring; conduct risk detection	NLP, anomaly detection, network analytics	Reduced false positives vs. pure rules; better detection of suspicious behavior	Privacy concerns; ethics of monitoring; risk of false negatives and regulatory scrutiny

4.3. AI in Market Risk Management

AI models are gaining ground in the market-risk literature as an approach to forecasting price dynamics, volatility and losses, and learning complex dependencies between macroeconomic variables and market variables on price dynamics. The use of machine learning techniques like gradient boosting or deep neural networks is presented as a way for modeling nonlinear relationships between risk factors that could enhance VaR and expected shortfall estimates in normal and stressed scenarios. AI-driven risk-measurement analysis is also covered as a tool to create stress scenarios and model different market scenarios, giving risk managers a broader array of options than parametric risk-measurement analysis (Giudici et al., 2024). Despite this, the increased complexity and opacity of AI models create concerns regarding interpretability and validation, especially in the context of regulatory capital, and highlight the need for proper model risk management.

4.4. AI in Credit Risk Management

Another big application of AI in the field of credit risk management is counterparty credit assessment and early warning systems. The use of credit scoring models based on artificial intelligence is mentioned, as they are proposed to integrate financial statement information, market-based indicators (like credit spread), behavioral data, and alternative data (like sentiment from news and supply chain data) to create more dynamic and granular measures of default risk. These models are said to be used for exposure limits, collateral requirements, and pricing of complicated credit products in investment banking. The benefits of earlier detection of credit-deterioration signals via AI are mentioned in terms of more proactive risk mitigation, including adjustment of exposures or renegotiation of terms. Concurrently, AI-driven credit models are also extremely sensitive to the quality of the data, and can introduce biases if the training data are themselves based on historical disparities or gaps, which is an ethical and regulatory concern for both fairness and transparency.

4.5. AI in Operational and Compliance Risk Management

AI has been shown to be valuable in analyzing unstructured data and detecting anomalies in operational and compliance risk functions in the literature. AI models can be used to analyses transactions, system logs, and incident reports to detect suspicious patterns that may suggest internal

issues or external attacks, such as fraud detection, cyber-incident detection, and process-anomaly detection. In compliance risk, AI-powered monitoring systems are said to leverage natural language processing and anomaly detection to analyze trade activity, emails, transcripts of chats and voice communications for indications of market abuse, insider trading, or misconduct. These systems are mentioned in terms of lowering the number of false positives produced by purely rule-based systems, so that compliance teams can focus on the highest-risk alerts. There are, however, privacy and ethical issues to consider, and models trained predominantly on historical instances could not detect uncommon or emerging types of misconduct.

4.6. Governance, Model Risk, and Ethical Considerations

AI has become a staple in investment banks' risk management practices, which is raising several new challenges for model governance, validation, and ethics (Habbal et al., 2024). One of the primary issues with AI systems is that they can encounter model risk when their models are complex, hard to understand, or are trained on biased or incomplete data, resulting in inaccurate risk measurements or biased outcomes. Effective governance has been defined in the literature as having clearly defined roles and responsibilities for model development, model validation, and continuous monitoring as well as adequate documentation, performance testing, and back-testing. The importance of explainability and transparency is particularly highlighted in models used for capital calculations, client decisions, and regulatory reporting, leading to an increased interest in explainable AI techniques. Ethical issues involve delivering non-discriminatory results, protecting sensitive information and maintaining human oversight and control over AI systems.

The categories of financial model risk, as generally recognized in the past in the light of supervisory guidance, for statistical, econometric models are different from the categories of model risk which are more important when it comes to deploying machine-learning or deep-learning systems. The model risk management approach, as laid out in the Basel Committee on Banking Supervision's model risk management guidance, the Office of the Comptroller of the Currency's model risk management guidance, and the Bank of England's principles for banks [SS1/23] (Bank of England, 2024), focus on conceptual soundness, ongoing monitoring, and outcomes analysis of models whose form and parameters are often given a priori. These elements become even more complicated when machine-learning or deep-learning systems are employed: When the model learns to capture the interactions of features, such as high-dimensional features, which are nonlinear, the conceptual soundness review becomes more difficult, and when the model is used to monitor continuously, it must also track data and concept drift gradual or sudden changes in the statistical relationship between the features used in the model and the output being predicted that can occur between the scheduled revalidation dates and unknowingly affect the model's performance; and when the model is employed

in the outcomes analysis, the population or the behavior it was designed to monitor may change over time, such as when a credit decision is made based on AI analysis, which can be a feedback loop that also alters the population or the behavior the model is dealing with.

The literature has identified additional governance aspects that go beyond the traditional model risk management, if applied to AI systems in an investment-bank risk function. According to Giudici et al. (2024), flexible models can perform poorly when experiencing out-of-distribution and adversarial inputs, in addition to being validated against past outcomes. Explainability requirements differ depending on the use case and the audience: Model developers and validators generally require mechanistic explanations, at the level of features, to debug and challenge a model; customers, auditors and supervisors may require outcome-level explanations, justifying a specific adverse decision, more often (BCBS, 2025). Accountability and human oversight are discussed as being in the form of designated owners for each phase of development in the AI lifecycle, as well as established escalation and override procedures which would allow for the identification of a human decision maker to be held accountable for any risk decision made on the basis of AI, as opposed to the blurring of accountability across a development team, a vendor and an automated system.

Data governance involves access management, data quality controls and lineage tracking of the often large and diverse data sets, such as alternative and unstructured data, used to train AI risk models. Bias and fairness assessment includes testing the model's output on various protected groups or sensitive groups according to multiple and sometimes conflicting definitions of fairness, and reporting the trade-offs made when fairness constraints are imposed (Bahlool et al., 2026; based on the existing literature on credit scoring). Issues with granular customer and counterparty data and the potential for AI systems to remember or leak moments of their training data all lead to privacy concerns. Robustness is the resistance of model results to minor changes in model inputs, such as attacks to avoid fraud or to avoid intrusion detection systems. Post deployment monitoring, documentation and auditability are reported to be dependent on automated drift and performance dashboards, model inventories and version control of training data and features, as well as validation results, to be able to recreate why a particular decision was made. Lastly, lifecycle governance is not a one-off approval gate prior to deployment but is part of the development, validation, deployment, monitoring, periodic re-validation, and eventual retirement/replacement of a model. The literature indicates that AI is not just a new addition to an existing model risk management checklist, but rather that some of the elements of the checklist, such as validation, monitoring, and explainability, must be rethought when it comes to models that learn and can drift.

4.7. Implications of the Findings

The review states that AI can play a meaningful role in enhancing the effectiveness of investment

banks' risk management in a proper governance framework. AI tools are mentioned as being used to strengthen risk measurement capabilities, such as early identification of irregularities in transactions and communications and more proactive risk mitigation (Jagirdar et al., 2025). The advantages of AI are also reliant on the quality of the data, the effectiveness of model risk management, and ethical and regulatory concerns. From the literature, it seems that investment banks that can successfully incorporate AI technologies into their risk management processes and overcome these challenges are more likely to achieve competitive benefits and greater resilience, while investment banks that overlook governance and ethical aspects could be subjected to increased financial, reputational, and regulatory risks.

This review does not include a primary survey of risk managers or a systematic inventory of the use of AI models at specific banks nor measured data on the allocation of AI related technology budgets across risk categories. For these reasons, no percentage, counts, or other estimates of numbers are included in this report on these matters. The applications most discussed in the literature reviewed, based on the frequency and emphasis of the discussion, are related to market risk and credit risk, which play a central role in the regulatory capital and profitability of the banks, and compliance-related surveillance, and to a lesser degree, dedicated tooling for operational risk. In a similar way, the most cited advantages of the sources reviewed are enhanced predictive power and better detection of misconduct and anomalies, followed by gains in efficiency, data integration and regulatory reporting. These observations, which represent the relative emphasis of the literature and are not empirical measurements, serve to describe the relative emphasis.

4.8. Critical Analysis: Conflicting Findings and Limits of the Evidence Base

The previous sections have reviewed the literature on how AI applications are described, this section explores where there are disagreements in the evidence, where AI-based approaches don't consistently outperform traditional approaches, and the implications of the findings for the application of AI in regulated banking practice. The first source of tension has to do with the ubiquitous assertion that the ability to make better predictions in AI models must be traded for less interpretability. In one recent systematic review of AI credit scoring, the performance difference between inherently interpretable models (like logistic regression) and black-box models is often small in studies which explicitly compare them, and when quantitative comparisons have been reported between the same type of models with the same data preprocessing, the difference is typically a few percentage points of AUC or smaller (Bahlool et al., 2026). This implies, at least in the realm of credit scoring, that the concept of a "cost" of nontransparent models, which is stressed in much of the previous literature, is often claimed but not proven, and that the decision to use transparent versus opaque models is as much a function of institutional appetites and regulatory policy as it is of an inherent technical cost.

A second, more regular finding, which was not related to explainability, is that having stricter fairness requirements leads to lower predictive accuracy; in the literature on credit scoring, multiple studies reviewed by Bahlool et al. (2026) report that more stringent parity requirements result in greater degradation than less stringent requirements do. This means that not every trade-off for AI is the same, and there are trade-offs that are assumed in the literature and those that are measured and confirmed.

While the reviewed papers provide many examples of AI models outperforming traditional models, there are fewer papers that discuss specific circumstances under which AI models fail to outperform traditional models. This is another limitation of the reviewed papers that may be subject to publication bias. The limitations reported are grouped around three situations: (a) small, noisy, highly imbalanced datasets (e.g., rare fraud or default events), where flexible models can be prone to overfitting the noise when simpler models are more effective; (b) structural breaks or regime changes (e.g., sudden shifts in market conditions), where models trained on historical data may generalize poorly until retrained, which is also noted in the market-risk literature reviewed in Section 4.3 (Giudici et al., 2024); and (c) situations where the prediction problem is already simplified by extensive feature engineering and dimensionality reduction, and more complex models are unlikely to improve performance (Bahlool et al., 2026). These conditions suggest that AI risk models should not be presumed to be better than traditional models in general, and that a comparison test against well-specified traditional models should still be performed, not as a formality but as a requirement for validation.

In addition to the above limitations, translating research models into regulated banking environments also adds some more constraints which are less visible in the academic literature reviewed herein, as most of the literature is based on public or semi-synthetic benchmark datasets and not on any proprietary banking data. This leaves a gap between the model performance reported in one study and the performance that can be expected once the model has been retrained on the individual bank's own, and typically less well-documented and less rich, data; some of the sources examined recognize this gap without being able to rectify it, as proprietary bank data is, by definition, unavailable to academic researchers, as described in Bahlool et al., 2026). Academic evaluation metrics like AUC or F1 score are not necessarily the metrics capital or conduct-risk rules regulate for the deployed systems, for example, however, such as expected shortfall or false negative rates for suspicious activity reporting. This makes published results more difficult to directly transfer to the deployed system. The benefits outlined in Sections 4.3 to 4.5 must be interpreted as representative of the potential of AI-based approaches in the data and regulatory context studied in the literature, and not guaranteed in actual deployment in any given investment bank's data and regulatory context.

4.9. Regulatory and Ethical Expectations: A Comparative Overview

Since financial institutions have a wide variety of model-risk and supervisory needs, this section offers a comparative analysis, using authoritative regulatory and standard-setting sources and not general AI literature, across the following areas cited in Section 4.6 as being central to AI governance: transparency, explainability, validation, accountability, data-protection, and human-decision-making. Supervisors cater to transparency and explainability expectations in two related ways. First, general model-risk-management guidance, such as the U.S. Federal Reserve and OCC supervisory framework, which reflects the experience of conventional statistical models, was written with conventional statistical approaches in mind and can be applied to AI models, but is unclear who the explanation would need to benefit—the model developer, model validator, business user, or supervisor—and to what level of technical detail (BCBS, 2025). Second, consumer-protection statutes such as the Equal Credit Opportunity Act (ECOA) and the Fair Credit Reporting Act (FCRA) impose a separate and specific transparency requirement, namely to provide the adversely affected individual with a rationale for the adverse credit decision within a specified period, which in practice limits the opacity of a credit-scoring model, even if it is highly accurate overall over a specified population of customers (as summarized in Bahlool et al., 2026).

There are banking-specific and cross-sectoral instruments to address validation and accountability expectations. Since 2022, the Basel Committee on Banking Supervision has been looking at AI and machine learning, including the allocation of responsibility and accountability for AI-driven decision making among model developers, validators, business users, and independent auditors, in the context of explainability and governance structures; and followed this up in December 2024 with more detailed supervisory insights on the regulation of AI in the financial sector (BCBS, 2024). As AI usage continues to grow, the Financial Stability Board's (FSB) 2024 evaluation also notes model risk, data quality and governance issues as key vulnerabilities that current financial policy frameworks may not yet fully cover, closely related to the external third-party and vendor dependency of many investment banks in developing their own models.

Requirements for data protection and human decision making are closely related to issues of fairness and bias as outlined in Section 4.6. The emerging AI Act in the European Union, places obligations on risk management, data governance, technical documentation, transparency, and human oversight on sensitive credit-scoring and creditworthiness-assessment systems that will be deemed to be at high risk, and also imposes restrictions on the use of sensitive data and mandates a basis for the use of automated decision making that significantly affects an individual (as summarised in Bahlool et al., 2026). These are similar expectations that are expressed in other jurisdictions, although the forms and levels of formality and maturity of oversight vary, suggesting an overall convergence in international expectations of explainability, documented human oversight, and contestability of AI-driven decisions in the regulated lending and risk functions. The principle that unites the regulatory

sources reviewed herein is that AI can inform risk decisions, but that it is not acceptable to leave the responsibility for a decision to the AI itself, and that the institution should be able to explain on demand why an AI produced a particular decision and who is accountable for taking action based on that decision.

5. Conclusion

5.1. Summary of Key Themes

The literature has been reviewed on the use of artificial intelligence in risk management specifically in context of investment banking and in relation to market risk, credit risk, operational risk and compliance risk. The review shows that AI is indeed useful for identifying, measuring, and monitoring risks in a meaningful way, as it is able to process large, unstructured and complex datasets, detect nonlinear relationships, and reveal a hidden pattern that a traditional model might not be able to find. AI is mentioned in the sources reviewed as a way of improving short-term forecasting, intraday VaR estimation, and stress-scenario generation in market risk, more granular and dynamic counterparty assessment in credit risk, and fraud detection and conduct surveillance in operational and compliance risk. Meanwhile, the literature points to the potential introduction of fresh challenges with AI implementation, such as increased model complexity, quality issues, explainability concerns, and ethical and regulatory challenges, that necessitate robust governance and oversight.

5.2. Conclusions

The literature examined is consistent with the notion that AI can be a valuable complement to quantitative models and human expertise for investment-banking risk management, rather than a replacement. In a suitable governance regime, AI-powered tools may seem to be able to improve the speed and quality of banks' access to risk data, their ability to act swiftly against new risks, and the effectiveness of banks' risk management resources. However, if AI models are used without proper validation, documentation, and human oversight, they can generate additional potential model risks and affect trust in risk measures and risk judgment. The report concludes that adopting AI in risk management must be done with a balanced approach of innovation and prudence, and that the decisions on risk management should be taken by humans.

5.3. Recommendations

The five recommendations below do not constitute best-practice statements but rather follow logically from the evidence and gaps identified in the review, and are supported by the section where the evidence is discussed. First, investment banks need to establish, document and demonstrate AI strategies that clearly weigh the use of AI against risk appetite, business goals and regulatory

requirements, in response to the supervisors' expectation to see institutions have clear evidence of ownership and decision-making over their AI use when asked, as well as the FSB's (2024) observation that governance gaps are the main vulnerability associated with the use of AI in financial services. Second, financial institutions should enhance their model risk management framework for AI systems, including by implementing out-of-distribution and adversarial stress testing, as well as adopting regular monitoring of model drift instead of point-in-time validation, and by documenting enough to be able to explain individual model decisions, in line with the explainability-audience distinctions discussed in BCBS (2025) supervisory analysis. Third, institutions are encouraged to foster close collaboration between risk managers, data scientists, IT specialists and compliance officers, with the rationale for this that it is clear from Sections 4.3 to 4.5 that AI value-adds only when technical model building is combined with domain-specific risk judgment and regulatory awareness, and not just a data-science exercise. Fourth, targeted training and capacity building should be done to enable risk professionals to understand how to interpret the results from models, including simple explainability methods, and to exercise informed override authority as identified in Section 4.6 and Section 4.9 as a common theme in both the academic and the supervisory literature. Last but not least, banks should have clear, measurable, testable fairness and ethical principles for the use of AI in risk management, including notions of bias, fairness, transparency and data privacy, and document the trade-off to be made between fairness and accuracy, consistent with the evidence provided in Section 4.8 that there is a measurable and consistent accuracy cost from applying fairness constraints that institutions should be able to justify and disclose, but not leave implicit.

Acknowledgments

Not Applicable

Author contributions

Ratul Hasan was solely responsible for conceptualization, methodology, formal analysis, investigation, data curation, and writing original draft preparation and writing review and editing. The author has read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data availability

No new datasets were generated for this study. The analysis is based on secondary literature cited in the reference list.

Declarations

Conflict of interest

The author declares that there is no conflict of interest.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Additional information

Received: 17 July 2026

Accepted: 18 September 2026

Published online: 22 October 2026

Publisher's Note

Wisdom Academic Press Ltd. remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Ashta, A., & Herrmann, H. (2021). Artificial intelligence and fintech: An overview of opportunities and risks for banking, investments, and microfinance. *Strategic Change*, *30*(3), 211–222. https://doi.org/10.1002/jsc.2404

[2] Attipalli, A., Mamidala, J. V., Kurma, J., Enokkaren, S., Kendyala, R., & Bitkuri, V. (2025). *Systematic review of artificial intelligence techniques for enhancing financial reporting and regulatory compliance* (SSRN Working Paper). https://doi.org/10.2139/ssrn.5741343

[3] Bahlool, R., Hewahi, N., & Elmedany, W. (2026). Performance, fairness, and explainability in AI-based credit scoring: A systematic literature review. *Journal of Risk and Financial Management*, *19*(2), Article 104. https://doi.org/10.3390/jrfm19020104

[4] Bank of England. (2024). *SS1/23: Model risk management principles for banks*. Prudential Regulation Authority. https://www.bankofengland.co.uk/prudential-regulation/publication/2023/may/model-risk-management-principles-for-banks

[5] Basel Committee on Banking Supervision. (2024). *Regulating AI in the financial sector* (FSI Insights No. 63). Bank for International Settlements. https://www.bis.org/fsi/publ/insights63.htm

[6] Basel Committee on Banking Supervision. (2025). *Managing explanations: How regulators can address AI explainability* (FSI Occasional Paper No. 24). Bank for International Settlements. https://www.bis.org/fsi/fsipapers24.pdf

[7] Financial Stability Board. (2024). *The financial stability implications of artificial intelligence*. https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/

[8] Giudici, P., Centurelli, M., & Turchetta, S. (2024). Artificial intelligence risk measurement. *Expert Systems with Applications*, *235*, Article 121220. https://doi.org/10.1016/j.eswa.2023.121220

[9] Gregory, J. (2012). *Counterparty credit risk and credit value adjustment* (2nd ed.). Wiley. https://doi.org/10.1002/9781118673638

[10] Habbal, A., Ali, M. K., & Abuzaraida, M. A. (2024). Artificial intelligence trust, risk and security management (AI TRiSM): Frameworks, applications, challenges and future research directions.

Expert Systems with Applications, *240*, Article 122442. https://doi.org/10.1016/j.eswa.2023.122442

[11] Ho, T. S., & Lee, S. B. (2004). *The Oxford guide to financial modeling*. Oxford University Press. https://doi.org/10.1093/oso/9780195169621.001.0001

[12] Jagirdar, S. S., Gupta, P. K., Jain, S., Vibhu, G., & Sharma, M. (2025). AI-driven innovations in banking analytics and regulatory compliance. In *Advances in Computational Intelligence and Robotics* (pp. 15–38). IGI Global. https://doi.org/10.4018/979-8-3373-0209-6.ch002

[13] Lee, J. (2020). Access to finance for artificial intelligence regulation in the financial services industry. *European Business Organization Law Review*, *21*(4), 731–757. https://doi.org/10.1007/s40804-020-00200-0

[14] Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. *Journal of Management Analytics*, *6*(1), 1–29. https://doi.org/10.1080/23270012.2019.1570365

[15] Martín-Martín, A., Costas, R., Van Leeuwen, T., & López-Cózar, E. D. (2018). *Evidence of open access of scientific publications in Google Scholar: A large-scale analysis*. https://doi.org/10.31235/osf.io/k54uv

[16] Namdar, K., Wang, P.-C., Raju, T., Zheng, S., Li, F., & Khan, S. T. (2025). *Anti-money laundering machine learning pipelines: A technical analysis on identifying high-risk bank clients with supervised learning*. arXiv. https://arxiv.org/abs/2509.09127

[17] Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3), Article 160. https://doi.org/10.1007/s42979-021-00592-x

[18] Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*, *15*(3), 398–405. https://doi.org/10.1111/nhs.12048

[19] Wacker, J. G. (1998). A definition of theory: Research guidelines for different theory-building research methods in operations management. *Journal of Operations Management*, *16*(4), 361–385. https://doi.org/10.1016/S0272-6963(98)00019-9